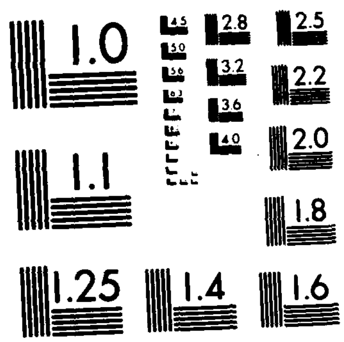
AD-A134 903 BEHAVIORAL CORRELATES OF SYSTTM OPERATIONAL READINESS 1/1
(SOR): SUMMARY OF W..(U) NAVY PERSONNEL RESEARCH AND
DEVELOPMENT CENTER SAN DIEGO CA D MEISTER OCT 83
UUCLASSIFIED NPRDC-SR-84-5 F/G 5/11 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

NPRDC SR 84-5

OCTOBER 1983

AD-A134 903

BEHAVIORAL CORRELATES OF SYSTEM OPERATIONAL READINESS (SOR): SUMMARY OF WORKSHOP PROCEEDINGS

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED

NAVY PERSONNEL RESEARCH
AND
DEVELOPMENT CENTER
San Diego, California 92152

NPRDC Special Report 84-5 October 1983

BEHAVIORAL CORRELATES OF SYSTEM OPERATIONAL READINESS (SOR): SUMMARY OF WORKSHOP PROCEEDINGS

D. Meister

Reviewed by
R. E. Blanchard

Released by
J. W. Renard
Commanding Officer

Navy Personnel Research and Development Center
San Diego, California 92152

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER NPRDC SR 84-5	2. GOVT ACCESSION NO. AD-A134 3	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) BEHAVIORAL CORRELATES OF SYSTEM OPERATIONAL READINESS (SOR): SUMMARY OF WORKSHOP PROCEEDINGS		5. TYPE OF REPORT & PERIOD COVERED Final Report Oct 1981-Sep 1982 6. PERFORMING ORG. REPORT NUMBER 17-82-18
7. AUTHOR(s) D. Meister		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Navy Personnel Research and Development Center San Diego, California 92152		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62763N, RF63-522-001-014-03.01
11. CONTROLLING OFFICE NAME AND ADDRESS Navy Personnel Research and Development Center San Diego, California 92152		12. REPORT DATE October 1983 13. NUMBER OF PAGES 80
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes a 2-day conference called to explore the methodology required to develop a behavioral model of system operational readiness (SOR). Participants discussed (1) the behavioral variables that should be included in the model, (2) the system level measures that should be included, (3) the mechanisms that permit behavioral variables to affect system output and SOR, (4) the kind of data needed to exercise a behavioral model of SOR, and (5) the way personnel policies

affect the behavioral variables. The major points made by participants have been extracted and form the body of this report, along with written responses to the five topics above. It appears that the development of a behavioral SOR model is possible, although expensive and heavily dependent on the availability of operational (fleet) data).

FOREWORD

The workshop described in this report was funded under RF63-522-001-014-03.01 (Definition and Relation of System Readiness Components) and was sponsored by the Chief of Naval Operations (OP-115). The workshop was convened to explore the needs, issues, and methodology required for development of a behavioral model of system operational readiness. The participants included eight distinguished researchers selected from the behavioral and operations research scientific community and three senior staff members of NAVPERSRANDCEN.

J. W. RENARD
Commanding Officer

J. W. TWEEDDALE
Technical Director

SUMMARY

Problem

Since policy makers are unable to anticipate how changes in personnel policy will affect fleet operational readiness, it is difficult to make effective use of personnel. The purpose of the overall project is to develop a methodology that will enable the Navy to relate personnel policies to personnel performance and personnel performance to system operational readiness (SOR). One method that has been suggested is the development of a behavioral model of SOR.

Objective

The purpose of the effort described in this report was to explore the methodology required to develop such a behavioral model and to consider alternative approaches to the solution of the problem.

Approach

Eight specialists from the behavioral and other disciplines participated in a 2-day seminar, together with staff and management personnel from the Navy Personnel Research and Development Center, to discuss the following questions:

1. The set of behavioral parameters/variables that should be included in the desired model and the way they should be defined and measured.

2. The system level measures that should be included as being descriptive of the output of a ship/squadron fleet and representative of their SOR.

3. How to conceptualize the mechanisms that permit the behavioral parameters/variables to exercise an effect on system output and SOR.

4. The kind of data needed to exercise such a model and where such data could be located.

5. The way personnel policies affect the behavioral parameters/variables.

Results

The major points made by workshop participants have been extracted and summarized and form the body of this report. The written responses to the five questions are provided in the appendix.

Conclusions

It appears that development of a behavioral model of SOR is possible, although it will be expensive and heavily dependent on the availability of operational (fleet) data of a behavioral nature. If such a model is developed, it should consist of a hierarchy of submodels "nested" into each other at three levels: individual performance, team/subsystem performance, and total ship performance.

Recommendations

Workshop participants recommended that:

1. A "front-end" study of user requirements be performed.
2. A study be performed to determine what required operational data are available.
3. A specification be developed to describe the model anticipated.
4. Development of a submodel on a single "high payoff" parameter be initiated.

CONTENTS

INTRODUCTION

Problem

For some years, the personnel community has sought a technique for predicting how a particular policy action or man-machine design approach would affect some index or measure of more ultimate system performance, such as fleet system operational readiness (SOR). Without such a technique, there is no way to predict how personnel policy changes (e.g., training program alterations, selection standards) would affect fleet SOR. The need is fundamental and continues to demand attention because, until it is satisfied, the Navy has no basis (other than opinion) for determining whether personnel resources are being used effectively.

It is important that personnel resource investment decisions be based on accurate estimates of how each available option would contribute to fleet SOR. The problem may be labeled with a different descriptor, but it concerns the continuing requirement for a technology that can be used to predict how various policy positions and/or human factors design options would impact (in terms of value, payoff) on more ultimate measures of system operational performance (e.g., operability, availability, readiness, maintainability, reliability, etc.). The three services have made considerable efforts over the years to determine how behavioral outputs in man-machine systems relate to a set of more ultimate system performance output indices. However, since the problem has not yet been solved, the concepts developed by previous researchers apparently have been inadequate in some as yet undetermined manner: Little has been accomplished in this area during the last 10 years.

Background

An indispensable ingredient in the actual SOR of the U.S. fleet (regardless of how much it is ignored in theoretical studies) is personnel performance--both enlisted and officer. If personnel are not available, if they cannot do their job properly, if they are not motivated, the probability of successfully accomplishing operational missions becomes unacceptably low.

During the late 1960s, the Chief of Naval Operations (CNO) (OP-98) and the Bureau of Naval Personnel (Pers A-2) sponsored research to develop quantitative approaches for modeling the man-machine system and for predicting the impact of design-specific performance outputs on overall measures of system performance. One such output was the technique for establishing personnel performance standards (TEPPS), which was tested on Navy combat information center (CIC) and sonar systems but was never implemented (Smith, Westland, & Blanchard, 1969). During the 1970s, considerable effort was devoted to the concept of human reliability modeling and data through the Navy's ADP043-13X improved human engineering program. Various approaches to man-machine modeling were explored, along with requirements for a Navy human performance data base (Blanchard, 1973). During the same period, the Air Force sponsored several studies on modeling man-machine human performance data sources and other studies concerned with relating elemental levels of human performance to more ultimate measures of system output.

Previous studies of the problem of predicting SOR have encountered a number of serious difficulties that also affect the present project. The following problems are especially serious:

1. The variables involved are multidimensional and highly complex. The severe difficulties in operationally (concretely) defining some of these variables are exacerbated with behavioral variables like motivation and leadership.

2. There is no agreement as to what constitutes SOR. The definitions that are available are multidimensional and described in a very subjective manner.

3. Empirical data to describe relationships among system variables and between system variables and system output (i.e., SOR) are lacking and very difficult to procure.

4. A serious deficiency of most SOR concepts is inadequate recognition of the behavioral elements in the SOR model.

5. Very few SOR models view readiness in performance terms. Rather, the viewpoint of most SOR models concerning readiness is that, if the required numbers and ratings of personnel are available and these have formal school training, a ship (from the standpoint of personnel) is operationally ready. This viewpoint stems from an equipment orientation; that is, if an equipment is functioning and meets its specifications, it is maximally ready and there are no lesser degrees of readiness for that equipment. Such a viewpoint is not applicable to personnel who have a variety of performance outputs, ranging from those that are completely ineffective to those that exceed criterion. The fact that a man with a particular rating for which he is presumably trained is in place on a ship does not mean that the man will function optimally in combat. The behavioral concept of SOR relates specifically to a prediction of how personnel will perform when required to do so operationally (i.e., in the ultimate analysis, in combat).

Richards, Eirich, and Geisler (1980) discussed in some depth some of the difficulties of studying SOR. In reading the following quotations from their report, it must be recognized that what they say is in the context of logistical considerations.

> That the readiness management problem has remained elusive for so long is a sign that numerous and complex difficulties are involved. These difficulties include:
>
> 1. Inconsistent definitions of readiness both within and across the services. This phenomenon is the result of a number of circumstances. First, readiness is the concern of almost all defense management functions. But each function has different interests in mind in its decision making. Logisticians, for example, view readiness from a perspective quite different from those of operational commanders or research and development engineers. Second, there is more difficulty in specifying "output" measures for some combat units and weapon systems than others. In an effort to reconcile definitions with measures, some definitions get stated in terms which do not reflect a potential wartime output with respect to unit or weapon system missions. Third, a topic of considerable debate is whether some threat should be explicitly incorporated in the definition of readiness or whether some other term better connotes the capability to wage war against a specific enemy. . . .

2. Lack of quantifiable output measures for all mission and weapon system types. Many of the measures currently employed are static measures such as operationally ready rates or mission-capable status rates. These measures are justifiable only when a unit or weapon system is operating in peacetime at a wartime activity level. . . . Another type of measure, the C-rating of the Force/Unit Status and Identity Report, is useful to operational commanders; but the subjective content of such measures and the limited scale of measurement present problems in developing the linkages between resources and readiness. . . .

3. Myriad factors involved in the interrelationships which affect readiness. To those newly initiated in the intricacies of the logistics support of a weapon system, the number of variables and the complexity of the relationships between them may seem overwhelming. This renders an adequate understanding of the effect of the . . . process on . . . readiness a difficult problem in information processing and analysis. . . .

4. Excessive cost of conducting readiness exercises. Readiness exercises potentially offer one of the best methods for assessing actual readiness, for it is only during such exercises that most units are stressed to their wartime capability. Readiness exercises, however, are very expensive to conduct and the documentation of exercises . . . is often given subordinate priority to the demonstration of primary mission capability. Also, readiness exercises do not always adequately simulate wartime environments. (pp. 1-3 through 1-6)

From a computerized DIALOG search of citations describing SOR, only 12 studies were found that dealt wholly or largely with which can be termed "personnel readiness." These studies are described below.

1. In 1965-1966, the Navy Personnel Research Laboratory (NPRL) in Washington, D.C. performed at least two studies on SOR: Gray (1965) dealt with a literature search and development of a general approach to the problem of personnel contributions to SOR, and Gray (1966) investigated defense establishment actions in the personnel readiness area. The approach taken by the latter study was in accordance with the four "C" categories defining SOR: C-1, fully combat ready; C-2, combat ready; C-3, marginally combat ready; and C-4, not combat ready. Research recommendations were made for developing a predictor score based on performance in operational readiness inspections (ORIs).

2. Subsequent to these studies, Dozier and Black (1967), NPRL, correlated ORI and ORE (operational readiness exercise) scoring data with on-board personnel counts for 17 destroyers (DD) and 18 auxiliary ships. The variables studied were number of personnel on board in various ratings, percent of allowance on board, percent of stability, etc. Results showed, as one might expect, little correlation between the personnel numbers aboard ship and ORI/ORE performance.

3. However, results of a study by Lockman, Stoloff, Manheimer, Hardgrave, and Story (1970) suggest that personnel variables may be related to ORI. They used multiple regression to relate ORI scores to training, equipment, supply, and personnel (e.g., number

on board by rate or number of weeks of training) variables. Three performance factors or groups of similar measures (control procedures, casualty control procedures, and ASW tactical communications) were found from a principal components factor analysis of the correlations among the 21 ORI subsections. For each of the three ORI factors, all resource variables (including personnel and training) were related to their component ORI subsections. The resulting multiple correlations were very high (about .80) for control procedures and ASW tactical communications, and substantial (.55) for casualty control procedures. The percent variance accounted for by the various personnel variables was substantial (e.g., the number of lieutenants on board accounted for 20% of the variance in control procedures, the number of engineering group leaders accounted for 52% of casualty control procedures, etc.). However, results show only that there is a relationship, not how the relationship is implemented.

4. In 1976, the Navy Personnel Research and Development Center (NAVPERSRANDCEN) conducted a study relating human resource management (HRM) and SOR, as measured by Navy refresher training (REFTRA) (Mumford, 1976). Although HRM cannot be considered a measure of personnel readiness, it can be conceptualized as involving behavioral variables. It would, therefore, be of interest to determine if a relationship exists between HRM and SOR. REFTRA scores, which are known to be somewhat subjective because of the manner in which REFTRA is conducted, were obtained for 34 Pacific Fleet ships whose personnel had been administered HRM. Correlation coefficients computed between the HRM indices and two types of REFTRA scores (full REFTRA and final battle problem scores) were highly positive. Over half were significant at the .01 level for the full REFTRA, but these were not sustained for the final battle problem. Hence, the results of this study are only suggestive. Correlation coefficients can suggest a relationship but, without a model, the mechanisms for the relationship cannot be ascertained.

5. Some interest has been shown in relating organizational variables to SOR. LaRocco and Jones (1977), in a theoretical study, emphasize the need to consider the ship as a total system. Olmstead, Baranick, and Elder (1978) found high correlations between organizational process dimensions and Army combat effectiveness, as measured in a computer-assisted map maneuver system (CAMMS) simulation (a table game). The dimensions of the organizational process (e.g., information acquisition, providing information and intelligence, planning, decision making, coordination, communication, clarity of roles, etc.) were rated by three levels of players. The CAMMS exercises also provided an effectiveness score. Although it was noted that the study met few of the conditions that must be met if CAMMS is to reflect actual combat effectiveness accurately, seven of the nine process dimensions were highly correlated with bridge command group effectiveness. Again, there appears to be some sort of relationship between personnel-related variables and effectiveness (which can be conceived of as a form of SOR), but how the relationship occurs is unknown.

6. Wagner (1979) performed a study relating leadership style to aircraft material readiness. Data were gathered from 19 naval squadrons whose mission was to fly and maintain the A7E aircraft. Leaders in the four key positions in each squadron were given Fleishman's leadership opinion questionnaire (LOQ) and Fiedler's least preferred co-worker (LPC) questionnaire. Data on 16 readiness variables were collected for each squadron over a 6-month period. These variables were grouped by factor analysis into five factors: equipment, operations, personnel, mission, and material readiness. Aircraft material readiness was defined to include aircraft availability, flight operations, manpower utilization, mission capability, and maintenance procedures. Correlations between leadership style and the criterion variables were tested by bivariate correlation, canonical

correlation, and multiple regression analysis. Bivariate correlations provided five significant (.01 level) correlations between leadership-style variables and SOR, multiple regression provided three significant correlations, and canonical correlation provided four significant results out of 60 analyses.

7. Holtz and Phillips (1969) attempted to develop models that reflect the impact of personnel turnover and military occupational specialty (MOS) mismatch on the readiness of Army units. The Army readiness reporting system, like that of the Navy, has four categories: fully ready, substantially ready, marginally ready, and not ready. These are reported by the unit commander with a substantial infusion of subjectivity. The basis of the personnel and training classification is a personnel count. For personnel, the system requires a report on strength and MOS match; for training, it requires the percentage of personnel turnover by grade and rank and the percent of personnel stabilized for the past 3 months. Two models were developed: a strength-turbulence indicator that took the form of an equation ($100 \times \frac{\text{effective strength}}{\text{authorized strength}}$) and an MOS indicator. The concept underlying the latter is that relative values can be assigned to MOSs in a unit, which would then permit degrading unit personnel counts, with all the disadvantages of this assumption.

8. Niehl and Sorenson (1968) developed a model (SIMPO-I) determining the qualitative impact of personnel policies. The model attacks through simulation problems related to assignment of personnel to jobs or job categories. It is stochastic; the basic population from which simulated individuals are randomly sampled is the multivariate normal, although nonrandom sampling resulting in nonnormal distributions may also be simulated. Studies performed with this model deal with the allocation of personnel, as described below:

> At the start of the simulation, individuals in a sample are optimally assigned over c jobs on the basis of criterion performance. During the first t_i months, each of the N individuals is examined for possible loss from the system. This loss may be a function of an individual's estimated job performance, simulated events occurring within the system, a random process which determines that near to p percent of the sample will be lost, or some combination of these variables. . . .
>
> For each individual remaining in the system, new performance measures appropriate to his job assignment and the length of time he has spent in the system are simulated. To replace men lost to the system, new random numbers are generated and transformed to expected values for an inexperienced population. Assignment to different job categories is performed such that expected performance of the new sample is optimal and job quotas reduced by loss are restored to their original values.
>
> At the end of each point in time, t_i, the effectiveness of the system is evaluated. The evaluation may be as simple as computing the average measure of job performance for the different job categories, or it could involve a fairly complex function of the performance of crew members where a weapon system is involved. Simulation then proceeds from time t_i to time t_{i+1} by again testing observations for the N individuals for loss of retention in the system

> and by generating new observations to represent enough inexperienced personnel to fill the losses. . . .
>
> With this type of simulation, quality of predicted performance in the system can be examined as the proportion of experienced personnel in the system increases. In addition, the rate or change in rate at which men are lost from the system can be related to system performance.

From the description of the SIMPO-I program, it is unclear whether it has been validated or what the system effectiveness measures referred to consist of. As described, it is unlikely that SIMPO-I would satisfy the requirement of relating behavioral parameters to SOR.

9. The Center for Naval Analyses performed two regression analysis studies correlating various personnel variables to mission capability. Horowitz (1977) found that two variables that consistently affected casualty report (CASREP) downtime were average crew pay grade and equipment complexity. Crew size and length of service (LOS) were influential for some ratings. These conclusions refuted those of a previous study by Berning (1971), who studied the effect of ship overhaul on its subsequent material condition. Personnel variables were included in the analysis as control variables. The results of the analysis with respect to the personnel variables were disappointing, with the associated coefficients of correlation being insignificant and often in the wrong direction.

Analysis of the studies described above and of SOR studies in general lead to the following conclusions:

1. Very few studies of behavioral performance influences on SOR have been performed. Although the list above may not be completely inclusive, it describes the majority of such studies.

2. Most SOR researchers and modelers recognize that personnel variables are related to SOR, but they have generally emphasized the physical aspects of SOR (e.g., reliability, maintenance, logistics, etc.). Hardly any effective research on the behavioral inputs to SOR has been performed.

3. Where personnel are considered in SOR studies and models, in general only the availability aspect of personnel is dealt with, as reflected in personnel counts, the match of required (SMD) manning with actual manning, etc. Presumably this is so because it is easier to "count noses" than to try to measure personnel performance.

4. Nothing vaguely indicative of a model of behavioral inputs to SOR has yet been developed.

Purpose

In view of the preceding conclusions, it seemed useful to attempt to develop a truly dynamic model of behavioral influences on SOR. Until such a model is developed and tested, the direction future research on this topic should take will remain obscure.

APPROACH

Before extensive additional work is performed, the components of system readiness need to be better defined, and a conceptual framework to enhance understanding of the problem established. Immediate objectives require further conceptual definition of the problem.

Figure 1 illustrates the approach taken by NAVPERSRANDCEN. Component A refers to the various policies that govern allocation (or treatment) of personnel resources; Component B, the elemental human behaviors and equipment components that compose the man-machine system; Component C, the rational combination of those elements into a man-machine system whose performance, cost, development time, and so forth, are shaped by Component A; and Component D, the system output (state) indices or measures that are distributed on a continuum of ultimacy to the subject system's established mission performance requirements (e.g., destroying an enemy submarine at a specific depth, detecting an airborne target at a particular range, or processing information at a particular rate). The primary problem is relating the effects of Component A on the outputs of Component D, which can be done only by defining, quantifying, and providing a useful output metric or metrics for Component C, the man-machine behavioral components of the systems.

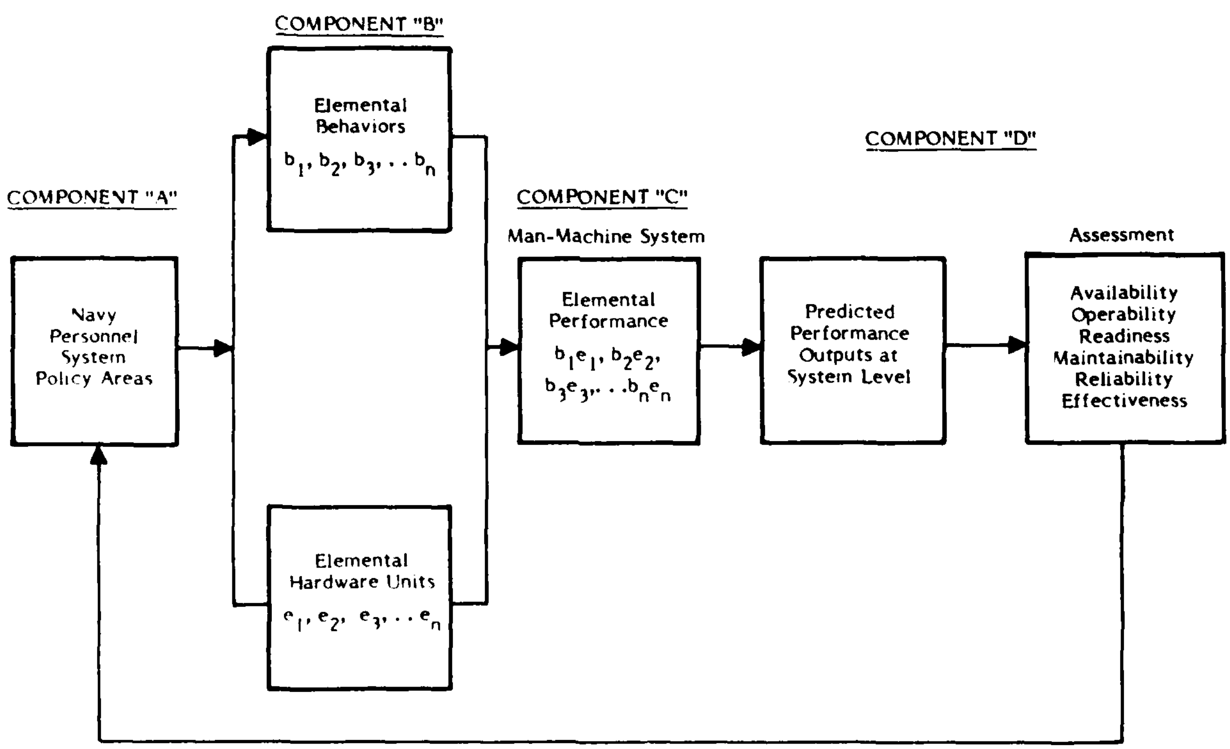


Figure 1. Components of man-machine system output prediction problem.

To develop ideas on how to develop a behavioral model of SOR, NAVPERSRANDCEN sponsored a workshop on SOR behavioral correlates, which was held in San Diego on 22-23 June 1982. Eight distinguished researchers from the behavioral and operations research sciences participated in the workshop, along with members of the NAVPERSRANDCEN management and staff:

o Dr. Robert E. Blanchard, NAVPERSRANDCEN.

o Dr. Alphonse Chapanis, Johns Hopkins University.

o Captain J. F. Kelly, Jr., then Commanding Officer, NAVPERSRANDCEN.

o Dr. David Meister, NAVPERSRANDCEN.

o Dr. Robert Mills, Human Factors Division, Wright-Patterson Air Force Base (submitted a paper but could not attend).

o Dr. Frederick A. Muckler, Canyon Research Group.

o Dr. Bernard Rotsker, Center for Naval Analyses.

o Dr. Mark Sanders, California State University, Northridge.

o Dr. Arthur I. Siegel, Applied Psychological Services.

o Dr. H. W. Sinaiko, Smithsonian Institute.

o Dr. Robert Williges, Virginia Polytechnic Institute.

Prior to the meeting, workshop participants were asked to think about the following questions, which are fundamental to the development of a behavioral model of SOR, and attempt to answer them by writing short papers. Participants were not constrained by the need to supply supporting evidence for any of their hypotheses (although such evidence would be desirable, of course).

1. What set of behavioral (and nonbehavioral) parameters/variables would you include in the desired model and how would you define and measure these?

2. What system-level measures would you include as being descriptive of the output of a ship/squadron/fleet and as representative of their SOR?

3. How would you conceptualize the mechanisms that permit the behavioral parameters/variables to exercise an effect on system output and SOR? In other words, how does subsystem behavior link to SOR?

4. What kind of data would one need to exercise such a model and where might one find these data?

5. How do personnel policies affect the behavioral parameters/variables? Or, how does one determine from an examination of a personnel policy/regulation which behavioral parameters/variables are affected by that policy/regulation? And how affected?

RESULTS

Workshop deliberations were recorded on magnetic tape and then transcribed. Because the resulting document was so long, a summary was made of the main points, organized around the following topics:

1. Varying concepts of readiness.
2. Readiness measurement methods.
3. Rationale for an SOR behavioral model.
4. Uses of an SOR behavioral model.
5. SOR model characteristics.
6. SOR model parameters and variables.
7. Data requirements.
8. Plan for SOR model development.

These topics are discussed in the remainder of this section. The papers submitted in response to the questions concerning the development of the behavioral model of SOR are provided in the appendix.

Varying Concepts of Readiness

For a research theme (SOR) that commands a great deal of interest, there is a serious lack of consensus as to its definition. This lack of consensus was reflected in the points of view expressed by workshop participants.

The disagreement narrows down to two opposed viewpoints. In the first, SOR is *defined, in a logistics sense, as reflecting that condition in which the reference system*, the ship, has (1) all its assigned equipment, which is functioning properly, and (2) all its required personnel in their assigned specialties, all of whom have graduated from appropriate schools and have attained some specified level of training at their jobs. The ship is then presumably ready to perform its assigned combat mission. If this definition is accepted, the present Department of Defense (DoD) method of calculating SOR (using the four "Cs" of the unit reporting system) is perfectly adequate.

The second, contrasting, concept of SOR is concerned with how well the ship will do once it has been plunged into combat. This concept suggests that defining SOR in logistical terms is insufficient because one wishes to predict the ship's ultimate performance. The ship is ready, but ready for what?

The distinction between the two concepts may reduce to a matter of chronology. The first or traditional concept of SOR applies to the ship when it is tied to a dock or is patrolling an assigned area. The second or effectiveness-related concept of SOR extends into the time period when the ship is engaged in combat. It was noted that the effectiveness concept of SOR encompasses the logistical one because, if a ship is to perform in accordance with its mission requirement, obviously it must be manned and its personnel must be adequately trained.

The question also arose as to whether there are degrees of readiness. Can there be a continuum of readiness so that the ship is entirely unready at one extreme and highly ready on the other? From a logistical standpoint, either the ship has the equipment and people needed and those equipment and people function, or the ship does not. If this is used as the basis of determining SOR, the SOR concept must be dichotomous. This is not the case with the performance-oriented concept of SOR because performance can be

represented as a probability (e.g., of winning a battle, of surviving an engagement, of error on an individual level). In fact, even though the C system of rating SOR is essentially dichotomous, the Navy treats the resultant values as if they were on continuum: It accepts the fact that a ship can be technically unready but only because of a minor deficiency that, if the need exists, can be remedied quickly. It was suggested then that the Navy accepts the logistical concept of SOR grudgingly because it lacks a more adequate way of defining and measuring SOR. There is, however, serious concern with the present concept, as demonstrated by the continuing efforts to research SOR. If the logistical concept of SOR is considered adequate, there is little point to conducting further research efforts such as that which led to this workshop.

It was also pointed out, most markedly by Captain Kelly, that one must distinguish between formal and informal methods of determining readiness. Although the formal C-rating of SOR is, of course, consulted in making an assignment of ships, the Navy's informal knowledge of both ships and men is equally important. For example, in a case where both are equipped to perform an AAW mission, the ship whose crew was better practiced in that mission, and whose commanding officer was more aggressive, etc. would be a better choice for that mission. This kind of information is not included in nor can it be derived from the formal SOR rating. It might be possible, of course, to add it to the formal rating.

In part, which concept of SOR one espouses depends on the use one has for the SOR measurement, the questions one would ask of SOR data, etc. This topic will be dealt with in greater detail later, but in this connection it is worth mentioning the frame of reference in which most of the SOR discussions took place. This was the recently concluded Falkland Island war between Great Britain and Argentina. The hypothetical situation addressed was that the conflict was between the United States and Argentina and the question to be answered by CNO was which ships to assign to the area. Assuming that CNO had a choice of ships (i.e., all were equally ready in a logistical sense), the criterion he would apply in making his choice would be that of anticipated performance effectiveness. Indeed, assuming that all ships are ready for sea, that is the only basis of selection; otherwise, one might as well choose randomly.

There was a difference of opinion among workshop discussants between those who felt it necessary to "nail down" a definition of SOR and those who felt that it had been adequately defined in terms of mission accomplishment of effectiveness. It was pointed out many times, notably by Dr. Rotsker, that different users may have different SOR concepts. This point will be addressed in greater detail later.

Whichever concept of SOR one adopts, one still must deal with the question of the elements in that concept. Dr. Muckler pointed out that the current unit reporting system has four variables on which the C rating is based: personnel readiness, equipment on hand, equipment readiness, and personnel training. Two of these four are personnel-related, which accepts the importance of the behavioral element. However, Dr. Muckler raised the question as to whether these are the necessary and sufficient set of variables. Should not leadership be included? Unit cohesion? Is training the correct term or should there be more concern with skills resulting from training? Dr. Muckler also pointed out that, in view of the fact that two of the four elements in the unit reporting system are behavioral, behavioral scientists are responsible for telling the Navy whether these behavioral elements are the right ones to measure, whether they are being measured properly, and whether they do in fact tell the Navy the personnel subsystem status of the force.

It was pointed out that, in general, NAVPERSRANDCEN views the problem of researching the behavioral correlates of SOR as simply a special case of the problem of relating subsystem outputs (e.g., personnel performance) to total system output (e.g., SOR, mission accomplishment). The total system, the ship, is composed of both equipment and personnel, and the Center wants to know how one part of the system, personnel, impacts upon or affects ship performance. Viewed in this way, the problem is fundamental to behavioral science, because the determination of the relationship between subsystem performance and system output underlies every area of behavioral science. It is also extremely complex because the number of personnel parameters and variables that could influence system output is very large.

The overall project was viewed in terms of four subareas, which must be addressed if a feasible research strategy is to be implemented.

1. The relationship between personnel performance and system outputs. In concrete terms, for example, how does the detection capability of a sonarman contribute to the effectiveness of the destroyer's attack on a submarine?

2. The relationship between personnel parameters/variables and personnel performance. The number of personnel parameters (e.g., leadership, training, selection, motivation, experience, etc.) may be reasonably small but each is composed of several variables that makes the total number of possible effects very large. Training, for example, would have a number of variables, such as length of course, instructor/student ratio, number of students, number of hours of practice on simulators, etc. How is each of these related to the sonarman's detection performance?

3. The relationship between personnel policies and personnel performance. In concrete terms, what is the effect on SOR of a reduction of 10 percent in training time for naval aviators and their proficiency as fighter pilots?

4. The relationship between policies and system outputs (SOR). How does the 10 percent reduction in training affect the ship's capability to perform its mission?

It is subarea 4, of course, that we are ultimately interested in, because an understanding of this relationship enables one to make meaningful policy tradeoffs. Generally, it was felt that each subarea had to be attacked individually; otherwise, the problem is too complex and becomes intractible.

Readiness Measurement Methods

It was pointed out previously that present formal readiness measurement methods emphasize counting of available equipment and personnel; if enough of these exist, the ship is ready. Since Dr. Muckler provided a detailed discussion of these measurement methods in his paper (see appendix), they are not described here.

It was pointed out that the C-rating of personnel is "just a snapshot" (i.e., it represents reality only at a moment in time). (This brings up the technical problem of appropriate sampling rate for UNITREP measurement.) It can be changed overnight by the addition or subtraction of one or two specialists. Captain Kelly referred to the practice of "cross decking," which is the process of transferring required personnel from one ship to another to remedy deficiencies. This is standard practice and is done "almost every day."

In addition to the formal readiness measurement, there is an informal one. It is recognized that two ships that have the same readiness rating are not necessarily the same in terms of capability to perform a particular mission. To quote: "And they (whoever 'they' are, but presumably the decision-making levels of the Navy) know who the good skippers are; the ships tend to have a history of goodness or a history of badness that's almost mystical. . . ."

This comment elicited the suggestion that it might be useful to try to determine the bases of the informal judgments and to quantify these through what has been termed "policy-capturing" studies. Officers are asked to indicate how various elements (e.g., training or morale) would affect field exercise performance or operational performance in general. The size of the correlation computed between an element and performance suggests its relative importance to that performance. Dr. Muckler, in his paper, described an Army study that tested company-level field performance of Army units in the United States and Europe (Department of the Army, 1977). Results showed that four variables--leadership, training, morale, and personnel turbulence--accounted for 90 percent of the variance of the criterion. This study had the following advantages: realistic battle scenarios including some live firing, all units totally manned, and all equipment available. The battle simulations were closely observed and the policy-capturing study was performed with the observers as subjects. The questions asked were what made the difference in performance between units and what contributed to unit effectiveness.

There are, of course, certain problems in performing policy-capturing studies. Since the judgments made by subjects in these studies are entirely subjective, the question arises as whether these judgments are related to performance, even if those making them insist that they are. To an extent, the outcome of such studies is a function of the particular variables included in the policy-capturing situation. One could of course argue that these judgments in themselves are of major interest. What do people think and how does this influence their decisions and actions?

Another difficulty is the significance of a particular variable that is given a high loading on or correlation with the criterion. For example, in one study mentioned by Dr. Muckler (Borman and Dunnette, 1974), 27 naval officers were involved in a policy-capturing study that produced 14 behavioral correlates of some significance. One variable that stimulated considerable discussion was "sack time." The question was whether or not sack time is a variable significant for positive or negative performance. Apparently the more time personnel spend sleeping while off duty, the poorer the criterion performance, perhaps because they are not studying their on-the-job training materials, etc. What does one do then with this behavioral correlate of sack time to improve performance? It was suggested that this is the kind of variable that might be used in a diagnostic manner to determine improvements to readiness. It was also pointed out that sack time per se is not directly related to the performance of any kind of job. Many behavioral correlates are not directly related to performance. This indicates that there are "intervening variables" in the system of operational readiness that are essentially invisible and that, therefore, must be conceptualized in some form of model.

Nevertheless, it was agreed that policy-capturing studies are one way of assessing the contribution of certain variables to SOR.

Rationale for an SOR Behavioral Model

Although most of the workshop participants appeared to accept the need to develop a behavioral model of SOR, a minority viewpoint held by Dr. Chapanis required some discussion. Dr. Chapanis championed a purely empirical measurement approach.

One objection to a model was that there are presumably so many independent and intervening variables that it would be extremely difficult to model them meaningfully. Any behavioral model that incorporated all possible variables would be excessively complex. The counter to this was that, precisely because of the many variables operative in the operational situation, it would be extremely difficult to measure them all in that situation. These variables are interactive and any single SOR index is almost always multidimensionally determined.

It appears that the two arguments would cancel each other out, especially since all agreed on the number and complexity of the behavioral variables affecting SOR. However, it was pointed out that (1) all the variables influencing SOR do not have equal impact and (2) various techniques, such as policy-capturing studies, could be used to determine or at least to hypothesize which variables are of greatest significance. By focusing on a reduced number of more significant variables, the model could be simplified.

The model could also be simplified by developing a series of submodels, each of which would deal with a smaller number of variables. Thus, there might be a training model, a motivation model, etc. Of course, this technique would present the further problem of integrating these submodels to create the overall model. Also, it assumes, theoretically, that the dimensional space is known to begin with.

The point was raised that, at some stage in developing the model, it would be necessary to collect empirical data, if only to validate the model after its completion. The question asked was whether it would be possible to rely solely on already existent data or to use subjective estimates made by experienced naval officers as a data source. The question of collecting data with which to exercise the model is indeed a serious one that will be addressed in greater depth later. For the moment, however, it was suggested that there are multiple data sources that might supply the needed data.

One question that arose was that, if one were concerned simply with predicting the effects of personnel performance on SOR, why bother with a model at all? Why would it not be possible to select a system output like success in fleet exercises and simply correlate this with some subsystem output like the number of targets detected, speed of sonar detection, or number of hours worked per week? The result would be something like a correlation matrix or a very large multiple regression equation. Since there would be no need to explain anything, there would be no need for a special model; the statistical technique used to measure the relationship between variables would, in a sense, define a model structure.

Of course, the question is not simply whether variable X is related (by some amount) to variable Y. For example, researchers know, or at least suspect very strongly, that personnel training is highly related to mission effectiveness. However, if this is all they know, there is not much they can do with the information. Moreover, any single variable is likely to be so complex in itself that it cannot be interpreted without some conceptual structure that attempts to explain its effect. Previously, attention had been drawn to the variable of "sack time," which in one policy-capturing study was suspected to be highly related to SOR. If one accepts such a relationship at face value, should one not encourage ship's personnel to spend more time awake?

There are generally accepted reasons for utilizing a model. For example, a model is useful when it is difficult or dangerous or otherwise impractical to measure variables in the operational situation. It has already been pointed out that it is difficult to achieve

the desired amount of control over variables in the operational environment. Moreover, it is unlikely that the Navy would provide researchers an opportunity to collect empirical data to the extent necessary. It is difficult for civilian researchers to get aboard ships and to have ship's personnel made available for measurement.

It was also pointed out that a model is necessary and is inherent in any empirical measurement scheme. One does not simply go out and measure without developing a strategy for doing so; that strategy is a model. One has, for example, to decide what independent and dependent variables are to be selected for measurement (because one cannot measure everything) and this implies a model. Even if one did not develop a model prior to measurement, such a model would be necessary as soon as one attempted to understand and explain any relationship found.

Actually, the conflict between modeling and empirical measurement is somewhat specious. The two are not mutually exclusive. One can visualize a situation in which an effort at empirical data gathering would run concurrently with and would feed into the development of a behavioral model. Certainly the model would help to structure the empirical measurement.

Uses of an SOR Behavioral Model

Participants felt that the uses to which a behavioral SOR model could be put was a critical question and spent a great deal of time discussing it. The reason is that, depending on whom one talks to, there are different uses for an SOR model, which imply different methods of conceptualizing and measuring SOR. There are at least four uses for any SOR model or measure:

1. Evaluation of ship SOR. Here one could determine the operational readiness of a particular ship, of ships of a particular class, or of a fleet.

2. Understanding the reason for SOR variations. Whatever the SOR is, it would be possible to examine the factors that produced this SOR.

3. Diagnosis of ship's problems. Assuming that SOR is unsatisfactory, what behavioral element(s) create the difficulty?

4. Determination of policy effects. The question here is how a given policy would affect SOR. Such information would permit one to trade off potential changes (e.g., a cut of __% in training duration or student throughput would result in a __% reduction in SOR).

Evaluation requires a single number or, at most, a profile of single numbers, perhaps one for each of the ship's subsystems or missions. Understanding mechanisms and problem diagnosis require multiple values. Determination of policy effects requires a family of curves something like a nomograph. Hence, the questions become progressively more complex. Whereas empirical relationships can produce an evaluation, only a model can provide the multiple curves needed to examine policy effects.

It is assumed that potential users of the SOR model exist at all governmental and Navy command levels. These users include the Secretary of the Navy and CNO and their staffs, the commander in chief (CINC) of a fleet, type commanders, ship captains, and Navy laboratories. Presumably the highest command level would be most interested in policy effects. CINCs and type commanders would be primarily concerned with the figure of merit their ship had received and the diagnosis of their ship's problems. Navy

laboratories, such as this Center, would be particularly interested in understanding the mechanisms causing behavioral elements to affect SOR. The varied uses may even have implications for the variables included in the model. For example, if one were to attempt to satisfy the needs of the Navy's training community, the model would heavily emphasize training variables.

It seems most unlikely that all user desires could be satisfied in the initial development of the model, if only because of financial and researcher constraints. Therefore, it seems necessary to perform what is known as a "front-end" analysis to determine specific user needs. Based on results, a particular user would be selected and his desired use would be emphasized.

It was pointed out that the original impetus for the project of which the workshop is an outgrowth was the sponsor's desire to determine policy effects on SOR. However, even here the <u>specific</u> user (e.g., particular Navy code) may have specific (not merely general) SOR needs that should be ascertained. In this connection, Dr. Rotsker argued vehemently that the project was unlikely to be successful unless it received strong support from a defined sponsor, a number of whom were identified.

The varied uses also imply a time factor. The traditional SOR C-rating is concerned with what can be termed <u>immediate time</u>; that is, the status of the ship at any specific moment--now. An interest in policy effects is of a more general and long-range nature.

It was pointed out also that it is important to distinguish between a model that simulates the functioning of an individual ship and one that deals with variables in a somewhat abstract form. These differences may be tied to the varied questions to be answered and the different interests of model users. Procuring the data to predict the effectiveness of a particular ship may well pose considerable difficulties. On the other hand, a more tractible question is to determine the relationship between variables and their effect on ships in general. Again, it was emphasized that, at least at the beginning of this project, it is unreasonable to think in terms of a model that will satisfy all users and all questions. Assuming that an investigation is performed of potential user desires, participants were warned that some user desires may be unreasonable because these users may not be aware of the constraints under which model development proceeds.

The front-end analysis would seek to develop a list of questions that the model might answer. Before a question could be included in this list, however, consideration would have to be given to time and money constraints. Some policy questions, such as the effect of changing political and cultural trends on personnel availability, are so vast as to be infeasible, even though some researchers are studying this question. There may be other questions for which purely empirical measurement will provide better answers than will a model. Such a question might be how reducing an ASVAB subtest score for admission to a school would affect school output and fleet performance.

Dr. Siegel pointed out that, at the level of very specific questions, the building of a model is not a difficult job.

> (Such a model would be) pretty good (sic) but you're going to now end up with something that has a limited specific purpose and may or may not ever be usable or used by anyone. As you get to a higher level (of question), you get into a much more difficult problem. . . . It's going to have less validity for a specific question but will have greater generality. . . . If you want to go to the very specific level, then what you are talking about is a set of decision aids. . . .

It was suggested that the model's architecture might be so developed as to allow a variety of outputs selectable by a particular user. It was agreed that the answer to this question would have to wait upon preliminary development of the model.

SOR Model Characteristics

In developing a simulation model, there are two alternative approaches, which were termed by discussants as the "top-down" and the "bottom-up" approaches. In discussing these approaches, it is necessary to recognize the various performance/measurement levels to be found in a ship system. The lowest level is that of the individual performer, followed by the team level (which may also be the subsystem level), the subsystem level, and the system level, the ship as a whole. Each level has a measurement set peculiar to it that feeds, so it is hypothesized, into the next higher level.

In the top-down approach, one starts by measuring the system-level SOR index, which is ultimately mission effectiveness, or the degree to which top-level system requirements are satisfied. This index is correlated with some measure of personnel performance or, more likely, with a variable related to personnel performance. This is because measures of personnel performance are usually found at the individual or lowest level and the top-down approach usually does not go down to that level (performance measures at the system or highest level tend to be combinations of personnel/equipment outputs). Thus, in the top-down approach, a training variable such as length of training of personnel aboard ship might be correlated directly with the SOR index.

The top-down approach has certain distinguishing characteristics. It is more likely to adopt a correlational stance (e.g., regression analysis) in the statistics it employs. By their nature, top-level variables are somewhat abstract because they tend to include more molar variables. The top-level approach does not require intervening or moderator variables, so its model architecture is somewhat simpler; however, for that very reason, the top-down approach tends to explain less than does the bottom-up approach. The problem of developing a taxonomy to describe variables at a series of levels within the system is avoided. There is no need to integrate variables at different system levels. Some of these characteristics, such as simplicity of model architecture, avoidance of integration requirements, and taxonomic development, are positive; however, others, such as the failure to explain as much as a bottom-up model, are negative.

The top-down approach has another advantage: Top-level decision makers are concerned about top-level variables only. For example, CNO has very little power to influence motivation aboard a specific ship because he is too far removed from that ship by intervening organizational levels. Thus, the top-down approach to SOR modeling would be more consonant with decision-maker problems and questions.

The top-down approach has a countervailing problem, however. If the model maker adds anything to the model (e.g., a change in variables or algorithms), the weighting of all model variables must change because, at the top-down approach, everything affects everything else. In a bottom-up approach, changes at lower system levels may not translate into a required change at a higher level; the model maker can change part of the model (at a lower level) and hold everything else constant.

The bottom-up approach also attempts to relate variables at the top system level to ultimate measures of SOR, but it does so by building a series of submodels, first at the individual operator/maintainer level, then at the team/subsystem level, and finally at the

total ship level. Manifestly, this process of building from "the ground up," as it were, creates problems of its own. One problem is integrating the submodels vertically. If one measures performance at the individual level (e.g., a single sonarman in a team of three), how does that performance contribute to the performance of the remainder of the team? How does the performance of the sonar team contribute to the performance of CIC as a whole? How does the performance of CIC contribute to a submarine kill? It is apparent that algorithms must be written not only to relate the variables at a single system level, but also to relate individual performance to team performance, team performance to subsystem performance, and subsystem performance to system performance.

The question arises as to whether these submodels at the lower levels are "nested" (i.e., "feed into" each other), so that the output of a model at the individual level becomes the input to the model at the team/subsystem level. It is possible that the submodels at the various system levels are in fact distinct models. No agreement was reached on this point.

Integrating lower-level models into the single-system-level model may also pose difficulties because the relationships among variables may change as one goes from one level to another. A variable that is highly significant at one level may become insignificant at a higher level because other factors "buffer" the variable and moderate its effect. For example, the length of training may be significant at the individual performance level but insignificant at the system level because other factors such as equipment malfunctions cancel out its effect. There may also be predictors at the system level that are unique to that level; these may interact with lower level variables that have been carried on from lower levels.

At lower system levels, subsystem measures of SOR are required that obviously will not be the ultimate mission effectiveness SOR index, since the mission is confined to ship level only. Hence, it will be necessary to develop intermediate SOR indices appropriate to the lower level one is modeling. Some suggestions along this line are provided in the appendix (Meister, p. A-1). Both the top-down and bottom-up approaches face this difficulty, but the top-down approach is more likely to avoid it since it is less likely to get down to the individual/team levels.

The bottom-up approach to modeling permits one to understand apparent discrepancies where, for example, an ASVAB score is important at one level but not at another. Because intervening and moderator variables can be used in the bottom-up approach, deviations can be traced. The top-down approach would not permit such an examination because its architecture does not contain the appropriate variables.

It was suggested that it might be possible to approach the problem by having <u>both</u> a top-down and bottom-up approach. If so, it would be necessary in some way to keep these approaches distinctly different. How this could be done was not discussed.

A continuing anxiety expressed by some discussants was that the top-down approach would be too insensitive to detail. This is true, of course, if the variables included in the top-down model are relatively molar.

In connection with the bottom-up approach, it was noted that developing a quantitative model with nested submodels at various system levels is a major task that could not be completed in less than 4 to 5 years. That is because, if the culmination of the model is the top system level, one must work one's way first through the various lower levels. On the other hand, if one starts with the top system level and the Navy's interest in the SOR question wanes, one has at least the top level that can be modeled in far shorter time.

In all of this discussion, it was recognized, both explicitly and implicitly, that the part of the SOR problem dealing with the relationship of policy to personnel variables would be postponed; the first order of business was to relate personnel variables to SOR.

SOR Model Parameters and Variables

To begin, one must distinguish between behavioral parameters and variables. A parameter is a molar, top-level variable that encompasses a number of more molecular variables. For example, the parameter of training includes such variables as length of curriculum, number of students throughput, number of courses, etc. There are relatively few behavioral parameters but the number of variables is very large. Table 1 provides a list (not necessarily completely comprehensive) of behavioral parameters and variables that may influence personnel performance and hence SOR.

In view of this large number of variables, it seems unreasonable to expect any single model to incorporate all of them, at least initially. It then becomes necessary to select a smaller number that can be hypothesized to have the greatest impact upon personnel performance and SOR. Such a selection should, of course, be made on the basis of empirical relationships demonstrated in the psychological literature. However, in view of the weakness of that literature, a considerable amount of expert subjectivity will also be involved.

One basic assumption made by workshop participants was that a variable is significant only if it is reflected in changes in personnel performance. For example, the number of training courses an individual takes has no significance if the job for which he was trained no longer exists or he is not given the opportunity to perform in that job. Such an assumption would seem to be obvious, but some researchers *are overly influenced* by theory rather than fact.

One difficulty in attempting to hypothesize the significance of these variables is that they include many intervening variables. In other words, variable X is related to SOR but only indirectly; that is, only as X affects an "invisible" variable Y. This is another reason why the top-down approach to model development seems somewhat simplistic.

Some variables, particularly those of a motivational nature, are particularly "tricky." It is difficult to study them directly, or to control and measure them as one would other independent variables because their relationship to performance is covert. They may be what one might term "contextual" variables in the sense that they influence the environment in which performance occurs, but not that performance directly. They may be akin to what has been dubbed "performance shaping factors." Moreover, it is difficult to infer the effect of motivational variables from an examination of performance, because performance is concurrently being influenced by many variables, some of which are much more obvious than is motivation.

Some relationships are highly visible, like the relationship between a ship's propulsion system and SOR. If the boilers fail to make steam, the ship will not leave the dock. Put another way, if any SOR parameter goes to zero, the SOR product is zero. Other relationships to SOR (both physical and behavioral), such as the relationship between sonar operator performance and SOR, are more tenuous, indirect, and difficult to recognize. If the sonar operator is performing at less than desirable efficiency, no one may even be aware of this because of (1) the number of system levels or nodes intervening between the sonarman and the actions resulting in ship performance, and (2) the number of factors interacting with the sonarman's performance (e.g., salinity of water, physical condition of sonar transducers, etc.).

Table 1

Behavioral Parameters and Variables that may Affect SOR

Parameter	Variable
Selection	Acceptance criteria for:
	Physical capability (strength/size)
	Psychomotor capability (vision/audition)
	General intelligence
	Special aptitudes (e.g., mechanical)
	Educational background (years of schooling)
	Age requirements
	Recruiting inducements:
	Enlistment bonuses
	School promises
	Job promises
	Test characteristics (ease/difficulty)
	Number of recruiters
	Increase in recruitment of women
Retention	Reenlistment bonuses
	Fringe benefits (PX, vacation time)
	Salary
	Promotion policy
	Sea/shore rotation policy
Training	Training content
	Course duration
	Number/type of training devices/simulators
	Instructor/student ratio
	Number of students throughput
	Number of training sites
	Class size
	Number of prerequisites
Human Factors	Human engineering:
	Man-machine interfaces (hardware/software)
	Internal components
	Procedures
	Technical manuals
	Job aids
	System characteristics (hardware/software)
	Maintenance concepts
	System organization
Assignment	Job priority
	Personnel availability
	Personal preference
	Sea/shore tour rotation
Leadership	Presence or absence of leadership
	Leadership style
Job design	Task characteristics
	Behavioral demands
Aptitude	Type
	Quantity
Motivation	Job satisfaction
Work experience	Relevancy
	Length/amount
	Recency
Working conditions	Physical
	Psychological

It was suggested an attempt could be made to measure not SOR itself but something related to SOR that would be easier to measure than SOR. What this intermediate criterion would be was never settled. Ultimately, however, one would have to measure SOR if only to validate the model.

The question was raised as to whether the model developer would have to include certain nonbehavioral variables in the model because such variables interact with the operator's performance. For example, would it be necessary to know what the ocean temperature in the sonar situation was because temperature did in part determine the operator's detection probability? Manifestly, however, since it is impossible to include everything in the model, a certain amount of error in the model measurement must be accepted.

One point of view expressed was that a major criterion for selecting behavioral variables for inclusion in the model should be whether the selection would be acceptable to fleet personnel. It was agreed that their opinion would serve as a major input to the selection process, but it could not be the sole criterion. Another viewpoint brought up was that only those variables for which empirical data could be secured should be selected; however, others objected on the ground that variables for which it would be easier to collect data might not in reality be those that strongly influence performance. The example given was that of the man who loses his wallet on the street at night but only searches for it under the street light because it is easier to search in that location.

The necessity for keeping the model relatively simple was emphasized, although whether this would be possible in view of all the behavioral elements in SOR was doubtful. In an effort to reduce the complexity of the modeling effort, it was suggested that researchers concentrate on the *relationship between certain variables and individual* personnel performance and work on the relationship between those variables and SOR at a later time. It was generally agreed that, however one begins, (1) the model must deal with personnel variables, personnel performance, and SOR, and (2) at least for the first stage of model development, the relationship between policy and all other variables would be ignored.

Leadership was one parameter that was given a great deal of consideration, particularly by Captain Kelly, largely because fleet personnel in policy-capturing studies (see appendix, Muckler) are firmly wedded to its importance, even though it is difficult to define. Leadership may have different meanings for different people at the various Navy levels; its meaning may be different for a ship captain than it is for CNO. During the discussion, it was sometimes defined as successful mission performance and, at other times, as the leader's consideration for his personnel. Captain Kelly pointed out that an impression of leadership quality was often gathered through very informal means; for example, from what one hears about a ship captain from other ship captains. Nevertheless, assuming that one can define leadership objectively and concretely, it would be possible to include it as a variable in the SOR model.

Data Requirements

A major problem encountered by anyone attempting to develop a model, especially a behavioral model, is the availability of data. It will be recalled that some workshop participants wished to establish the availability of data as a criterion for selecting model variables, but this concept was rejected. Nevertheless, there does not seem to be much point in creating a model that cannot be exercised because it lacks data.

One can look at data in at least two ways: (1) data requirements (i.e., data are needed to exercise the model), and (2) data sources (i.e., where such data can be found).

A list of data requirements is provided in the appendix (Meister). The major items (including all significant variables) are summarized below:

1. Measures of overall system (ship) and/or mission effectiveness, because this represents the system level SOR index.

2. Measures of individual/team and subsystem level performance effectiveness, because these also represent SOR indices at lower system levels.

3. Data on crew perceptions of leadership, motivation, and working conditions.

4. Data describing aptitude (e.g., ASVAB scores) of shipboard personnel.

5. Data relative to selection, training, and retention of ship personnel specifically and in the Navy as a whole.

6. Statistics about the reliability and availability of shipboard equipment, about maintenance actions and shipboard manning.

There are two data sources: (1) the behavioral literature, which represents, in most cases, experimental studies, and (2) the operational environment, which provides data collected in the relevant ship environment when personnel are performing their operational duties. The amount of data already available from either or both of these sources is not very extensive. In fact, there is comparatively little relevant data to be found in the behavioral literature and almost none has been collected from the operational environment. The data gathered from the literature are weak because the experimental situations in which it has been collected do not resemble the "real world." Although there is a popular presumption among fleet personnel that they handle masses of operational data because of the many reports that must be filed, examination of those data indicate that they also are not very relevant, despite having come from the "real world." For example, String and Orlansky (1981) evaluated the usefulness of currently available maintenance data (e.g., 66-1, 3M) for determination of maintenance training effectiveness and concluded that currently available maintenance data were not particularly useful for any purposes other than those for which the data system was developed.

It was apparent, therefore, to workshop participants that concurrent with (or rather as part of) the development of a behavioral model, there must be an intensive effort to evaluate the adequacy of operational data sources. Although researchers are relatively familiar with the behavioral literature, no one, including fleet personnel, is really aware of those operational data that might be of value to a behavioral project. Such an evaluation effort would try to determine (1) what useful (to the model) data have already been collected, (2) what operational data sources could be of value to model variables and to what extent, and (3) what changes, if any, in the data system would be required to make such data truly useful for behavioral purposes. It was suggested that, to assure desired control over the variables of interest, it might be advantageous to use the Navy's training and test simulators to test crew samples on shore. However, such simulators would have to be highly realistic if test results are to be generalized to the ship situation. The data secured in this way might satisfy the requirement for individual/team and subsystem level data, but it seems quite clear that, for ship system-level measures, it will be necessary to make use of actual ships performing maneuvers such as battle exercises.

It was Dr. Rotsker's opinion (substantiated afterward) that present fleet exercise data do not include personnel-related material.

The possibility of using subjective techniques such as critical incidents was also raised. Such measures could be secured on board ship with only slight interference with ongoing activities. In any event, the question of data availability and what to do about it must have an extremely high priority.

Plan for SOR Model Development

The culmination of the workshop was the development of a comprehensive plan to construct the SOR model. This plan was to describe the project from inception to conclusion. Naturally, the initial phases of the plan were much more detailed than the later ones.

The plan includes the following six steps:

1. Front-end user analysis.
2. Development of a model specification.
3. Evaluation of data sources.
4. Prototype model development.
5. OPNAV/fleet demonstration of the model.
6. Implementation of the model.

The steps are listed roughly in order of performance but it is anticipated that steps 2 and 3 would be performed concurrently. The most time and effort would be demanded by step 4, development of the prototype model. *The time frame contemplated for all 6 steps* is 5 years.

1. Front-end User Analysis. Because there are multiple users of SOR data, an attempt must be made to identify them to determine the kind of SOR information they need to answer their questions. Presumably only those potential users who are responsible for significant policy decisions involving SOR considerations would be approached because the policy parameter and the tradeoffs involved in policy decisions are the primary concern of this project. The user analysis would produce a list of questions that cognizant Navy organizations want to answer. These questions would then be examined to identify the relatively few that are feasible for research.

2. Development of a Model Specification. Assuming that the model would be developed by a contractor, it is necessary to provide a set of guidelines for the developer. This would be true even if model development were pursued in-house. It was agreed that the model specification would contain the following elements:

a. The purpose of the model, the conceptual framework (e.g., emphasis on personnel performance) in which it is to be constructed, the definition of the problem, and questions the model should aid in answering.

b. The variables to be incorporated in the model, including their definition, hopefully in operational terms.

c. The type of model desired (e.g., stochastic, deterministic, top-down, bottom-up) and its architecture (e.g., hierarchy of submodels, nesting).

d. Data requirements and data sources.

e. Model outputs and required analysis of outputs.

f. SOR criteria, their definition, means of developing them, and data relationships.

g. Compatibility of the model with other models such as LCOM.

It was noted that the development of the model specification involves much more than merely writing a document. Each of the model subsections listed above will require some analysis and investigation. Since results of the front-end user analysis will have to be assimilated into the specification, this step is a very considerable effort in its own right.

3. Evaluation of Data Sources. It was agreed that the availability of data, particularly data from operational fleet sources, will make or break the SOR model. The questions that need to be answered are (a) what data are needed, (b) where those data can be found, and (c) whether it is feasible to secure the data. Since the data requirements stem primarily from the questions the model is designed to answer, results from steps 1 and 2 must contribute to specification of that requirement. The characteristics of the data (e.g., the data must be continuous rather than point estimates) are important. It is essential also to determine how feasible it is to secure the required data. Conceivably, the researcher could identify and locate a required datum but find that administrative bottlenecks or lack of cooperation from a Navy agency would make the datum impossible to secure.

4. Prototype Model Development. The greatest amount of work is involved in constructing a prototype model. The term "prototype" implies that the model will be built in successive iterations of tryout and revision. However, in view of the hierarchical nature of the model, it is likely that the model will be constructed in the form of a number of submodels. It was agreed that, concurrently, some empirical data collection and/or crude experiments might be required. Part of the model development would also involve constructing a data base for the model. This might require developing a system/organization for collecting and analyzing the data. Model development will also involve at some point in time (later rather than sooner) the validation of the model, probably in terms of its submodels. Validation itself will be a difficult and extensive process. It is apparent, therefore, that this step would involve several coordinated lines of activity.

5. OPNAV/Fleet Demonstration. At this stage in the project, it is presumed that the model has been completed and validated. To secure acceptance from those who are intended to use the model, it is necessary to demonstrate what it can do (e.g., the answers it will provide). It was pointed out that the model will not solve problems but will supply data and inferences that will aid in problem solutions. At this point in time, it is impossible to provide details of how a demonstration of the model would be conducted, since extensive planning obviously would be required prior to setting up the demonstration. There is always considerable resistance in any organization to a novel idea, and the notion of behavioral correlates of SOR is still a novel idea although it has been circulated since the early 1950s.

6. Model Implementation. Assuming demonstrating the utility of the SOR model is successful, arrangements would have to be made for the Navy agencies most concerned to take over the model and its use as a normal part of their operation and for the Center to relinquish control over the model as an R&D project. It was noted that model

implementation will not be a simple thing to accomplish because a number of factors must be considered. The changeover from an R&D project to continuing use by nonspecialists may raise a set of problems that need to be anticipated and resolved.

CONCLUSIONS

1. Assuming that further work is pursued on the definition and measurement of the SOR behavioral correlates, the most promising approach appears to be the development of a behavioral model of SOR.

2. The development of such a model or, rather, a series of hierarchically organized submodels is an extremely complex and expensive process.

3. Before the development of such a model is initiated, assured financial support of the effort for a minimum of 5 years is needed.

RECOMMENDATIONS

Workshop participants recommended that:

1. A "front-end" study of user requirements be performed.
2. A study be performed to determine what required operational data are available.
3. A specification be developed to describe the model anticipated.
4. Development of a submodel on a single "high payoff" parameter be initiated.

REFERENCES

Berning, J. et al. Ship overhaul effectiveness (CNA Study 1157). Alexandria, VA: Center for Naval Analyses, June 1971.

Blanchard, R. E. Survey of Navy user needs for human reliability models and data (Rep 102-1). Santa Monica, CA: Behaviormetrics, December 1972.

Borman, W. C., & Dunnette, M. D. Selection of components to comprise a naval personnel status index (NPSI) and a strategy for investigating their relative importance (Rep NPSI-02). Minneapolis, MN: Personnel Decisions, Inc., March 1974.

Department of the Army. Women content in units force deployment test (MAX WAC). Army Research Institute, October 1977.

Dozier, N. E., & Black, R. L. An evaluation of selected criteria for measures of personnel readiness (Rep. WRM 68-1). Washington, DC: Navy Personnel Research Laboatory, August 1967.

Gray, K. W. Development of criteria for evaluation of personnel readiness afloat: preliminary analysis (Rep. 66-21). Washington, DC: Navy Personnel Research Laboratory, December 1965.

Gray, K. W. Procedures for evaluating personnel readiness afloat (Rep WRM 66-47). Washington, DC: Personnel Research Laboratory, June 1966.

Holtz, B. W., & Phillips, P. D. Improved models to measure Army personnel readiness (Tech. Paper RAC-TP-381). McLean, VA: Research Analysis Corporation, November 1969.

Horowitz, S. et al. Maintenance personnel effectiveness study (CNA Study 1090). Alexandria, VA: Center for Naval Analyses, May 1972.

LaRocco, J. M., & Jones, A. P. A systems approach to organizational functioning in the Navy (Rep. 77-25). San Diego: Naval Health Research Center, July 1977.

Lockman, R. E., Stoloff, P. H., Manheimer, B. H., Hardgrave, J. B., & Story, W. F. Navy readiness analysis system methodology study: Vol. I. Study methods and results (Study No. 27). Arlington, VA: Center for Naval Analyses, January 1970.

Mumford, S. J. Human resource management and operational readiness as measured by refresher training on Navy ships (NPRDC Tech. Rep. 76-32). San Diego: Navy Personnel Research and Development Center, February 1976. (AD-A022 372)

Niehl, E., & Sorenson, R. C. SIMPO-I entity model for determining the qualitative impact of personnel policies (Tech Res Note 193). Washington, DC: U.S. Army Behavioral Science Research Laboratory, January 1968.

Olmstead, J. A., Baranick, M. J., & Elder, B. L. Research on training for brigade command groups: Factors contributing to unit combat effectiveness (Rep. TR-78-A18). Alexandria, VA: Army Research Institute, June 1978.

Richards, L. D., Eirich, P. L., & Geisler, M. A. A concept for the management of readiness (Rep. LMI ML913). Washington, DC: Logistics Management Institute, January 1980.

Smith, R. L., Westland, R. A., & Blanchard, R. E. Technique for establishing personnel performance standards (TEPPS) technical manual (Rep. PTB-70-5) (Vols. I, II, and III). Washington, DC: Bureau of Naval Personnel, December 1969.

String, J., & Orlansky, J. Evaluating the effectiveness of maintenance training by using currently available maintenance data (Rep. IDA P-1573). Arlington, VA: Institute for Defense Analyses, August 1981.

Wagner, R. E. An analysis of the effects of leadership style on aircraft material readiness (Master's Thesis). Monterey, CA: Naval Postgraduate School, December 1979.

APPENDIX

COMMENTS ON QUESTIONS TO BE ADDRESSED IN WORKSHOP ON BEHAVIORAL CORRELATES OF SYSTEM OPERATIONAL READINESS (SOR)

COMMENTS ON QUESTIONS TO BE ADDRESSED IN WORKSHOP ON BEHAVIORAL CORRELATES OF SYSTEM OPERATIONAL READINESS (SOR)

Author: David Meister, NAVPERSRANDCEN

Parameters to be Included in a Behavioral Model of Operational Readiness

All personnel parameters that could be remotely related to system operational readiness (SOR) have been included in the model, even though the contribution of some of these parameters to SOR may be relatively slight. Some of these parameters, such as motivation and leadership, can be defined only tenuously and are measured only with great difficulty. They are included in the model primarily because the Navy wishes a complete model.

These personnel parameters are selection, retention, training, assignment (to a school or job), human factors (design of the man-machine interface), leadership, job design, aptitude, motivation, experience, and working conditions. Each of these parameters includes several variables that, for lack of space, have not been included.

Nonbehavioral parameters, considered as influencing the effect of the behavioral ones, include direct monetary cost (e.g., of implementing a personnel policy), loss of time (e.g., in maintenance, mission response time), and requirements for additional facilities (e.g., increased logistics support, additional training facilities).

System-Level Measures

SOR is considered as a prerequisite to efficiency in combat; hence, it is closely tied to certain performance outputs of the system (defined as the ship, the squadron or the fleet, depending on what one's terms of reference are). These outputs are, for example, the speed with which the ship can maneuver, the speed with which its weapons can be brought to bear, the total amount of firepower it can throw at an enemy. On a personnel level, SOR is related to such dimensions as the speed of detecting the enemy, skill in tracking him, etc.

The ultimate index of SOR is mission effectiveness, which is defined generally as performance of required ship/system responsibilities to specified requirements. This definition assumes that these responsibilities have been precisely and quantitatively defined (e.g., speed of a destroyer, 35 knots). This ultimate index is composed of or influenced by both personnel and equipment elements in the system.

Other indices represent contributions of subsystem factors to the ultimate system output (SOR index). Since the system has both personnel and equipment subsystems, two further SOR indices must exist:

1. Equipment availability, defined as the relationship between uptime and down-time.

2. Personnel availability, which is the percent of personnel aboard ship available for duty in appropriate ratings as required by the ship manning document (SMD).

If one asks what contributes to equipment availability, the answer is equipment reliability, which is defined by mean time between failures (MTBF).

Personnel performance is the medium through which personnel effects on mission effectiveness are accomplished. There are two indices of personnel performance effectiveness:

1. Operability, defined as mean time (or percent of total time) between errors in equipment operation.

2. Maintainability, defined as the relationship between required and actual mean time to repair.

Elements of a Behavioral Model (Mechanisms)

These elements include:

1. The personnel policy being implemented.
2. The performance of personnel in the affected ship/squadron/fleet system.
3. The performance of the equipment in the affected system.
4. Certain moderator variables, which also influence personnel performance.
5. Personnel parameters other than those under consideration at a particular time.
6. Nonbehavioral parameters such as cost and time.
7. SOR indices related to personnel and equipment.
8. SOR indices related to the system as a whole.

The model operates on the basis of certain assumptions:

1. Personnel policies affect SOR in two ways: directly and indirectly. Direct influence on SOR is exercised in the case of personnel availability as, for example, when a DoD directive cuts the number of recruits for the Navy by N%. Obviously, this would directly reduce the availability of personnel to the fleet (and therefore their availability aboard ship) without any intervening performance being required of them. More frequently, however, personnel policies influence mission effectiveness or ship/system availability indirectly through their efforts on personnel performance. For example, a personnel policy is promulgated whose effect is to lower intelligence standards for recruitment. Lowered intelligence means that personnel will find it more difficult to perform their jobs and the error rate will climb. Those recruits assigned after schooling to maintenance work will have an increased frequency of errors during maintenance. These errors will have two effects:

a. It will take longer to find the malfunctioning component, so that equipment is down longer, which reduces ship/system availability.

b. Some of the errors made by personnel will not be caught during the maintenance and will therefore lead to further malfunctions later on, thus reducing equipment reliability and again reducing ship/system availability.

2. The effect of a personnel policy on personnel performance will be influenced by one or more moderator variables. These moderator variables include:

a. The behavioral functions required by the job (e.g., detection, communication, navigation, etc.

b. The behavioral demands of the job (e.g., stress imposed).

c. The personnel requirements imposed by the job (e.g., intelligence, special aptitudes, skills and knowledges).

d. Idiosyncratic attitudes and motivation (e.g., job satisfaction, self-image, etc.).

Moderators influence functions in the following way. A personnel policy is implemented. Whether that policy will have a positive or negative effect (or any effect whatever) on the SOR of a particular job depends on the nature of the job being performed in the system under scrutiny. The assumption is that there must be a match between the dimensions inherent in the policy and those involved in the job. Assuming there is a match between policy and job dimensions, the specific effect of the policy in the performance of the job will be determined by whether the policy is responsive to the behavioral and personnel requirements of that job. Suppose, for example, that Job X has a certain task complexity that requires a certain aptitude. If the effect of a personnel policy is to reduce the availability of personnel with that aptitude, the job will be performed less well than previously and mission effectiveness will be reduced.

It should be noted that the effect of a personnel policy, when it is exercised through performance (i.e., indirectly), is very specific; it must hit a particular job or rating. That is because the effect is exercised through performance, and performance is always specific. When the effect is exercised directly, as in directly affecting personnel availability, the effect is much more widespread. A specific effect can also have a broad spectrum, as when it affects a number of jobs of a specific type, but such an indirect effect (through performance) is unlikely to be as wide in scope as a direct one.

Idiosyncratic attitudes and motivation also moderate the effect of a personnel policy. Assume that policy encourages the enlistment of recruits with higher than average intelligence. In a technology-intensive Navy, one would expect, as a consequence, improved personnel performance leading to improved SOR. However, if the new recruits are misassigned, so that they are given jobs to do that demand no intelligence and are in fact very boring, they will become dissatisfied and their performance will fall, thereby reducing SOR.

3. A personnel policy may exercise an effect concurrently on more than one behavioral parameter. So, for example, a policy lowering standards for recruits (thereby affecting the selection parameter) may have a primary effect on personnel availability by permitting a larger group to enter the Navy; it may also have a concurrent secondary (but perhaps not equal) effect on the training parameter because additional training facilities may have to be provided for these additional personnel. Any secondary effect like this one is assumed to have a lower probability of occurrence than a primary effect has; for example, the additional personnel may all be earmarked for immediate sea duty, thereby postponing the training impact.

4. The probability that a personnel policy will have an effect on SOR is finite and variable and rarely approaches 1.0. Because the effect is probabilistic, it is possible to apply probability mathematics to the analysis.

5. The impact of a policy is always ultimately on a job. A change in selection criteria may produce more or fewer recruits of a certain type but whatever SOR effect these recruits will have will be manifested in terms of job performance. Leadership or the absence of it is significant only in terms of how well necessary jobs get done in the presence or absence of leadership. The only possible exception to the above is a policy

that has a direct effect on personnel availability: If personnel are not available, the job cannot be done.

If we examine the factors that affect job performance most directly, we have the following list in order of assumed importance: job design, aptitude training, human factors, work experience, motivation, and physical work conditions. The factors that affect job performance indirectly and with lesser consequence are retention, leadership, and psychological work conditions. It is therefore necessary in analyzing the effect of a policy to ask which of the above the policy will affect. One must then ask how the effect on job performance will be produced (i.e., what is the casual chain of events?). Here the analysis will follow Figure 1. To determine the amount of effect, one must examine specific job requirements to determine whether the job imposes any special requirements for aptitude or training. The effect of a personnel policy on jobs of that type would be significantly greater. Two problems are inherent in this analytic process:

a. There are many jobs in the Navy and they have many dimensions.

b. The policy being analyzed will hardly ever refer to a specific job and the job-related dimensions will almost never be self-evident.

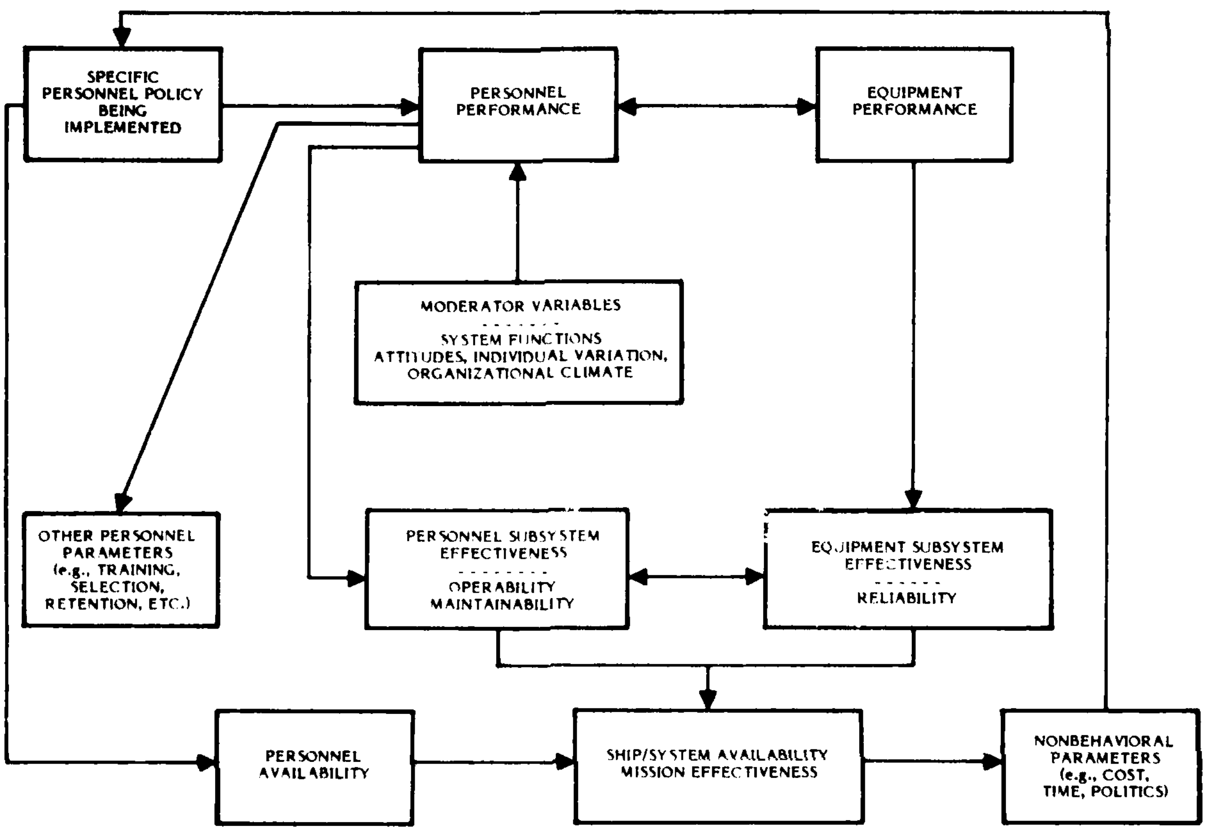


Figure 1. Model of the effect of personnel policies on operational readiness.

Data Requirements

What has been described is a purely conceptual framework for describing how certain effects are produced. Such a model is of no value in prediction because equations to which a quantitative data base can be applied are required for prediction. If these equations were available, it would be possible to transform them into algorithms and thus produce a computerized mathematical model as the end product of this project.

One must therefore begin by asking the question: Have the data already been gathered to quantify the concepts described previously or can they be gathered?

The kinds of data required are of the following types:

1. Using the individual ship as the system unit under consideration, but keeping in mind that with this proviso an N of at least 10-20 ships of varying classes and missions would be required, one would wish to gather quantitative data on the effectiveness with which each ship performs in various combat missions.

2. Also required are quantitative data on the performance effectiveness with which individuals and teams aboard ship in each of the major departments (e.g., propulsion, CIC) perform their jobs. These data (error and task accomplishment indices) would deal with both the operation and maintenance of ship equipment.

3. Subjective data with regard to crew perceptions of leadership, motivation, and physical and psychological working conditions aboard the ship.

4. Test data (e.g., ASVAB scores) relative to aptitude of the ship personnel performing as subjects.

5. Data (Navy statistics) relative to selection, training, and retention of Navy personnel in general and in particular relating to the personnel aboard the test ships and ships of the same class.

6. Statistics describing the reliability and availability of equipment aboard the test ships, including maintenance actions and the availability of pesonnel aboard these ships.

What is the availability of such data? Starting with the latter items first, it is certainly possible to secure data on equipment reliability and availability from 3M and other sources. The availability of ship personnel is largely a matter of counting noses relative to SMD manning. Presumably, Navy-wide statistics on selection, training, and retention variables should also be available, assuming cooperation of the cognizant agencies. The same applies to personal data describing personnel aboard the test ships, specifically ASVAB and indices such as supervisory ratings (although the latter are badly contaminated as representative of actual performance). Certainly personnel aboard the test ships could be administered questionnaires and other instruments for information about motivation, leadership, and working conditions. The human factors characteristics of equipment can be determined by inspection and reading of manuals. Job design information can be gathered by task observation and interviews. It is only with regard to objective data concerning the effectiveness of personnel performance and of ship mission effectiveness that one runs into difficulties, because actual personnel performance data are not routinely gathered and there is some question whether ship mission effectiveness data gathered by means of ORI fleet exercises are entirely objective.

The crux of the matter of predicting SOR is the relationship between the effectiveness of individual and team performance and measures of total ship performance, as represented by mission effectiveness measures gathered in ship/fleet exercises. Some of these fleet exercise data are routinely gathered but their usefulness to the determination of SOR is not known. The subsystem data (describing how well individuals and teams perform) will probably present the greatest difficulty because such data are not routinely collected.

The data are therefore theoretically available but whether they are in a useful form remains to be seen. First, it is necessary to determine the availability of these data from an access standpoint. In other words, will the Navy agencies responsible for the collection and analysis of these data release them for this investigation? Will type commanders make their ships and personnel available for observation, interviews, the administration of questionnaires, and the review of records? Almost certainly some of the required data are being collected, but the data may have to be somewhat modified for purposes of this project. Other parts of the required data are probably <u>not</u> being collected. For such items, it will be necessary for the investigators to mount a special data collection effort of their own. Whether the Navy Department would be willing to support (financially and in other ways) a data collection effort of some magnitude remains to be seen. The cost involved in securing individual and team performance data, for example, may be high.

The Navy Department may be willing to support the necessary data collection effort but it is wise to consider alternative ways of developing the data base for such a model (e.g., psychological scaling techniques making use of "expert" judgments).

Author: Robert G. Mills, Wright-Patterson Air Force Base

Although I will be unable to attend the workshop being arranged by the Navy, Dr. Meister has kindly consented to permit me to get my "two cents" worth in anyway. Also, I would like to add that Dr. Williges, who will be attending the workshop, and I have been working together on the systems research problem for a number of years. I am confident that he will be able to communicate my ideas as effectively as his own.

In order to initiate my commentary, I must preface it with the statement that the questions posed in the instructions for attendees are disturbingly foreign to me in the context of naval operations of which I know little or nothing about. I feel that, without such context, knowledge on the part of all those working the Navy SOR problem, and extensive systems analyses, the questions posed may be premature and virtually unanswerable. Therefore, I intend to briefly discuss an overall approach to the SOR problem, which has, as its objectives, the solidification and answering of the types of questions posed in the instructions. I would then like to summarize three projects being conducted here and with colleagues that have specific relevance to the systems research area, which, in my opinion, includes the Navy's SOR problem.

Assuming that the naval system (or subsystem) to be under observation is known and has been scoped (i.e., its boundaries defined), there are a number of representations of the object system. When developed successively, these system representations constitute an integrated process of system analysis that can lead to a better understanding of the system, identification of goals and objectives, identification of research and data collection issues, etc.

The Navy should be looking toward developing a systematic overall analytical/data collection approach that includes the application of interactive system analytical techniques, the use of efficient, economical experimental designs, and thorough pretesting of all experimental elements. The following is a listing of the types of system representations I would attempt to develop as part of a Navy SOR research program. (I must note that some of the descriptions below may seem vague and with good reason; the thinking is not solidified and only recently has evolved to a point where it can be put on paper.)

1. Real System Representation (R SYS REP)--The operational or conceptual system as it exists in the field in the case of the former, or on the drawing board in the case of the latter.

2. Structural System Representation (S SYS REP)--Consists of a structural decomposition of the system. The representation includes the hierarchy, system data flow, control flow (e.g., logic, policy, organization, authority, etc.), data structure, functions, and the linkage among these elements. Techniques for performing the structural decomposition process are available. They are manual and graphic in format, involving the development of block/network diagrams. We are currently using a technique called IDEF in a program for NORAD that has similar objectives to the SOR objectives. (System Development Corp. is our contractor on this project and they have a Navy group in San Diego.)

3. Goals/Objectives System Representation (GO SYS REP)--Includes all subgoals and subobjectives as well as SOR goals and objectives. Depending upon the level of the system's hierarchy, this representation could define the goals and objectives of single hardware and operator tasks. The distinction between a goal and an objective has been formalized in the work we are doing with Dr. Miller (see summarized projects below). A

goal may be any realizable or unrealizable final or intermediate state and may be qualitative. An objective is similar to a goal but is defined in the range of one or more performance measures.

4. Performance Requirements System Representation (PR SYS REP)--Defines the requirement levied on each system element and its linkages necessary to achieve goals and objectives.

5. Figure of Merit System Representation (FOM SYS REP)--The performance measurements representation, which includes defining measures of SOR, effectiveness, etc. down to lower level measures, depending on the levels of the system's hierarchy one is representing.

6. Data Sources System Representation (DS SYS REP)--Determines how (and from what sources) system and performance data will be obtained (e.g., operational exercises, literature, experimental, system simulation, etc.). Also, data availability and deficiencies are identified. At this point in the analytical process, the system performance questions of interest should be solidified.

7. Simulation/Model System Representation (SM SYS REP)--Includes real time or fast-time simulation and data collection during an operational or experimental (contrived) exercise.

I have identified seven types of system representations that can be developed successively in order to specify the system problems of interest and the means for solving these problems. By "successive development," I mean that the overall approach is a process of mapping one (or more) representation(s) into another. Thus, for example;

R SYS REP - - - - - - - - - - - - -> S SYS REP,
where - - - - - - - - - - - -> is read as "is mapped into or onto," and, for example,
(R SYS REP) X (S SYS REP) - - - - - - -> GO SYS REP,
where X is read "crossed with," and finally,
(R SYS REP) X (S SYS REP) X (GO SYS REP) X (PR SYS REP) - - ->
X (FOM SYS REP) X (DS SYS REP) - - - - -> SM SYS REP.

Additional Comments

1. It should be obvious that personnel issues and factors can be represented through the process. However, the analyst cannot ignore the existence of the rest of the system in which personnel are performing.

2. Each representation requires individual development but is part of a process of mapping one (or more) representation(s) into another.

3. There are other possible representations (e.g., alternate system configurations to be tested), and there could be further decomposition of any representation into a more specific one.

4. Finally, I have not commented on experimental design and data collection issues because I felt it more important, at this point, to discuss the system analysis problem. However, two of three projects summarized below should have some implications for the workshop and perhaps eventual application in the Navy's SOR program.

Summarization of AFAMRL/Contractor Projects

1. Formal Analytical Structure for Manned Systems is a project being conducted with Dr. Al Miller (OSU). The purpose of the project is to formalize and automate, in a user-assisted format, the structural decomposition process. We also intend to support the software with guidelines for its use. This has been a low-level effort (25K/yr) and we have finished 3 years. The basis for the effort lies in the use of set theory and assumes, in its simplest form, that a system is comprised of sets of objects, properties, and relations. In its final form, we hope to "bury" the set mathematics and terminology in the software in order that a user can work a problem at a terminal in a conversational mode. Thus, we are attempting to develop a formalized, automated structural decomposition technique that is disciplined and consistent (repeatable) and that will form the basis for creating all other types of system representations. If the money is there, we hope to begin a software development effort in FY87 using a portion of the NORAD system as a developing vehicle.

2. The User-assisted Automated Experimental (Test) Design Program (AED), Versions I & II, is a program by which a user can create full and fractional experimental designs interactively at a terminal. The program in its present form (Version I) can create 2-level designs for up to 15 factors at a time. Version II will be ready by the end of the fiscal year and has expanded Version I to include 2- and 3-level mixed designs, 3- and 5-level central composite designs, rules for creating optimum (in terms of information return) 2-level designs, and a storage facility for "calling up" previously created designs. While AED can be used to create one-shot designs, it can also be used to set up designs for a series of screening (efficient pilot) studies. In the latter case, one is faced with a large number of factors (greater than 5 or 6) that are present simultaneously in the system (simulation, etc.) and desires to collect only enough data to screen out those factors that are insignificant, leading up to a formal study involving the minimum set of significant factors. The output of AED is a list of observation vectors, each of which represents the level settings on each factor. The input is numbers of factors, levels, and trials along with some ancillary information. (System Development Corp. is our contractor on this project.)

3. Experimental Methodologies for Manned Systems Technology is a project being conducted with Dr. Robert Williges (VPI). The purpose of the project is to cross-reference systems research experimental methodological issues with available solutions (tools, techniques, etc.) and/or deficiencies (i.e., requirements for new tools, techniques, etc.). This is a low level effort (15K/yr) and has concluded its third year. Currently, we have developed a catalog performing the cross-referencing for a large number of issues covering the full range of the research process from study planning to data reporting. A draft copy of the last working paper will be available to the workshop via Dr. Williges. Planning includes beginning software development for a user-assisted automated capability to "call up" issues and techniques in FY83. We would like the automated capability to be fully self-contained; for example, a selected statistical technique could be executed as part of the program. Eventually, the program will be supported with guidelines and will be continually expanded by adding new issues and developing new techniques.

Author: Frederick A. Muckler, Canyon Research Group

Introduction

One Definition of SOR

It might be useful to at least start with one definition of system operational readiness (SOR) to establish some kind of initial framework. One convenient definition is from the Army (Department of the Army, 1979):

> The readiness of the Army as measured by its ability to man, equip, and train its forces and to mobilize, deploy, and sustain them as required to accomplish assigned missions.

For the purposes of this discussion, it is assumed that the term "Navy" can be substituted for "Army."

There are some practical constraints concerning the measurement of SOR that might well be mentioned immediately. They are:

1. This is truly a tri-service problem; any measurement approach must take into account the existing SOR reporting schemes that stress measurement comparability across services.

2. Any operational SOR measurement system must make minimum demands on the individuals and units required to prepare and report the SOR data.

3. The required number of dimensions must be as few as possible to express meaningful status estimates as well as rapid and simple indications for corrective action.

4. The financial costs of generating force readiness data must be held to a minimum; for each kind of data, it would be desirable to have tradeoff analyses between the value of the information and the cost of obtaining that information--in short, a cost and information effectiveness analysis.

For research purposes, there may be very large sets of behavioral dimensions for SOR. However, for an operational SOR reporting system, I would suspect that one would be very much bound by these four constraints.

One SOR Measurement Model

Following the definition given above, there is a sample SOR measurement model available. In notational form:

$$C_{1-5} = PR_C, EOH_C, ER_C, T_C \qquad (1)$$

where C_{1-5} = The overall unit rating as defined in Table 1

PR_C = Personnel readiness

EOH_C = Equipment on hand

ER_C = Equipment readiness

T_C = Training readiness

Table 1

Readiness Ratings

C-1	Combat ready, no deficiencies
C-2	Combat ready, minor deficiencies
C-3	Combat ready, major deficiencies
C-4	Not combat ready, major deficiencies
C-5	Not combat ready, service programmed
6	Cannot be determined (not a C-rating)

To measure these dimensions, measurement categories have been suggested, as shown in Table 2. It is interesting that two of the four basic dimensions concern behavioral variables in a broad sense: (1) personnel readiness and (2) training readiness.

Table 2

Dimensions of SOR Measurement

A. Personnel readiness data 1. MTOE required strength 2. Available strength percentage 3. MOS qualified strength percentage 4. Senior grade fill percentage 5. Personnel turnover percentage	PR_C
b. Equipment on hand data 1. Total reportable lines 2. Lines rated C-1 3. Lines rated C-2 4. Lines rated C-3 5. Lines rated C-4 6. Pacing item rating	EOH_C
C. Equipment readiness data 1. Reportable items 2. Equipment status (ES) percentage 3. Pacing item equipment readiness rating	ER_C
D. Training data 1. Weeks to be fully trained 2. Constraints a. Funds b. Equipment/material c. Training areas/facilities d. Fuel e. Ammunition f. Time g. Borrowed military manpower	T_C

Normally, in the case of a four-variable problem of this sort, it is customary to assume some model form for the combination of variables. One such linear example would be:

$$C_{1-5} = PR_C + EOH_C + ER_C + T_C \tag{2}$$

This immediately raises such questions as: Is this the right operator (addition) to combine variables? And what are the proper weightings that should be given to each variable? In fact, the current procedure avoids all these issues by requiring the overall rating to be equivalent to the lowest rating given any of the four variables. In short,

$$C_{1-5} = \min C\ (PR_C, EOH_C, ER_C, T_C) \tag{3}$$

which probably avoids many difficult mathematical and empirical questions. Whether or not this is the most valid and useful combinatorial rule is not known. And, indeed it might be of some interest to try alternative rules. The best advantage of the procedure given by (3) is that it minimizes the problem of combining a variety of measurement scales that are probably not mathematically or empirically compatible.

The logic of the rule, however, is conservative. The overall rating cannot be higher than the weakest component of the unit. Consider the following case:

Component	C-Rating
PR	1
EOH	1
ER	1
T	4
Overall	4

This describes a unit that has the people (in quantity at least) and the equipment specified for the mission but is poorly trained. And, it would take several weeks, if not more, to obtain fully trained status. This is surely not an inconceivable situation for a newly-formed unit. But, does the overall rating, C-4, represent a reasonable unit rating?

The Problem of Behavioral Correlates

It would probably be far easier to ignore the total SOR measurement problem and concentrate on just the behavioral dimensions. But it seems to me that one cannot, for the following reasons:

1. Whatever SOR reporting system is used, there is a need to specify the appropriate behavioral dimensions. In the example used here, the two dimensions are PR (personnel readiness) and T (training). Are these the "right" two dimensions? Should there be more? Or less? There is a great need for behavioral specialists to tell the people who develop SOR rating systems: (a) what personnel dimensions should be included, (b) how personnel dimensions should be measured, and (c) how they should be used.

2. Whatever behavioral dimensions are used probably cannot be too complex. There is the question of collecting the information, the question of the cost of the data, and the

critical question of proper interpretation of the data. We should take the responsibility of answering these questions; our "guesses" will be better than any others.

Questions Addressed

Behavioral Correlates of SOR

I know of no definitive data in the literature that defines the necessary and sufficient set of behavioral correlates for establishing SOR or the personnel subsystem part of the SOR. But, there have been some attempts to do so. Three past examples are shown below.

Example 1: Policy Capturing. In an earlier study using policy-capturing techniques, naval officers were asked to define the behavioral correlates that comprise SOR (Borman & Dunnette, 1974). The following 14 dimensions, which are said to possibly be the basis of a Naval Personnel Status Index (NPSI), were extracted.

1. (REENLIST)--Reenlistment rate.

2. (MAST)--Nonjudicial punishment rate.

3. (UAR)--Unauthorized absence rate.

4. (REENLIST-FT)--First tour reenlistment rate.

5. (TURNOVER)--Long-term stability of personnel.

6. (MANNING-NEC)--Manning level ratio describing the proportion of billets manned by qualified persons according to the NEC data system.

7. (PASS RATE)--Percentage of persons taking rating exams who pass.

8. (COURSES)--Rate of correspondence course participation.

9. (SACK TIME)--Percent of time spent sleeping while off duty.

10. (LEADERSHIP)--Perceived leadership effectiveness by crew members.

11. (DISCHARGE)--Discharges other than honorable (including reference to a special BUPERS code).

12. (MAINTENANCE)--Percentages of maintenance actions deferred due to insufficient manning or expertise (from 3M data system).

13. (SAT-PA)--Satisfaction with present assignment (as measured by confidential questionnaire to sample of officers and enlisted).

14. (ENLIST-FIT)--GCT + ARI + Education Level + Mean Semi-Annual Evaluation (all scores standardized) for all or for a sample of enlisted persons averaged across the ship.

A very important finding is that data on these 14 dimensions collapse factorially to five factors:

1. Fitness and readiness of individuals.
2. Discipline.
3. Crew member attitude toward officers and the Navy.
4. Free-time activities.
5. Manning level (corresponds to PR_C in the equations above).

Example 2: Unit Profile Analysis. A current system under development by the Ft. Hood Unit of the Army Research Institute has been "The Commander's Unit Analysis Profile" (1981). This is an attempt to develop a measurement tool by which the unit commander can assess the personnel readiness of his/her unit. This, in turn, should give some estimate of "mission readiness and operational effectiveness."

Currently, this profile has 88 items that have been tested on about 10,000 U.S. Army soldiers. From these data, the following 21 dimensions have been generated factorially:

1. Officer leadership.
2. NCO leadership.
3. Immediate supervisor leadership.
4. Leadership concern for soldier welfare.
5. Promotion policy.
6. Rewards and corrective action.
7. Leave and pass policies.
8. Quality of training.
9. Tools, equipment, and supplies.
10. Job satisfaction.
11. Freedom from harassment.
12. Military courtesy and discipline.
13. Race relations.
14. Unit cohesiveness.
15. Sports activities.
16. Social activities.
17. Freed from alcohol/drug-related problems.
18. Food.
19. Confidence in unit.
20. Morale.
21. Reinlistment potential.

A synopsis of the profile, including definitions of the dimensions and a description of the conditions under which the profile is used, is attached (see pp. A-19-A-22).

The main purpose of the profile is clearly diagnostic. The commander is looking in depth for the effectiveness and readiness of the unit. In turn, the commander will use that to correct deficiencies. In this case, none of these data go beyond the commander, but they could.

An important feature of the profile is that norms can be given against which to evaluate the particular unit score. This obviously requires a large data base--which apparently is available.

The studies on the profile to date suggest that all 21 dimensions do, in fact, discriminate at the company level. It would seem to be important to know the relative importance of all of these dimensions, or, in the sense of discriminant analysis, how much variance each dimension accounts for.

With this set and the previous one, there is the alarming possibility that a very large number of dimensions will be needed to account for the behavioral correlates of SOR; in these cases, either 14 or 21--or more. This situation may well be true, but it will make very difficult an operational measurement set that is easy to collect and relatively simple to use.

Example 3: Women in the Force. A final set of dimensions comes from the U. S. Army's MAX WAC field tests (Department of the Army, 1977), which tested company-level field performance with varying percentages of women in the units. The five dimensions uncovered here, with an estimate of their relative value, obtained by a policy-capturing method involving ratings of military officers who had observed the actual performance of the units, are listed below.

1. Leadership--30%.
2. Training--30%.
3. Morale--20%.
4. Personnel Turbulence--10%.
5. Percentage of women--5%.
6. Other--0%.

These, then, are three attempts to define the behavioral correlates of personnel subsystem operational readiness. They do not seem to have a great deal in common although the dimension of leadership does, for example, appear in all three.

System-level Measures

Of the five questions asked, this is by far the most difficult. There may be, in fact, hundreds of analytic studies that have attempted to define system-level measures for "ship/squadron/fleet" performance. The problem appears to be richness of measure sets, or, as one report put it--the chaos of dependent variables (Campbell, Bownas, Peterson, and Dunnette, 1974).

Some factors that might be considered are the following:

1. There appears to be a choice between the basic approach to measurement. The three examples just cited used judgmental rating scales. In short, "experts" rate according to some kind of scale and perhaps against some set of norms unit SOR.

2. For some, empirical approaches are fundamental. Short of wartime exercises, this must be either field exercises or simulations of operational performance. Speaking from personal experience, these tend to be enormously complex and expensive. In some cases, it may take months to get the data reduced (if they ever are).

A recently proposed partial alternative to field exercises has been the use of training devices where they exist (Hawley and Dawdy, 1981a, 1981b). These authors suggest that this notion is certainly technically feasible, but there are some questions about acceptance and cost.

3. If we return to equation (1), that in fact is a model of system level measurement. It says that we can measure operational effectiveness if we know the values (or some values) of personnel readiness, equipment on hand, equipment readiness, and training status. If we accept that as a satisfactory, first-approximation model, some questions that can be asked are:

a. Are these four variables necessary and sufficient? Should there be additional variables? For example, perhaps a separate dimension of command capability might be in order?

b. What are the relative weightings that could be given to these dimensions? Is the minimization algorithm shown in equation (3) appropriate?

c. What are the necessary and sufficient subdimensions needed for each of these four dimensions? What about, for example, "personnel turbulence" as a subvariable in personnel readiness? And, is "weeks to be fully trained" sufficient to estimate training status?

Right now, there are simply too many possible system measures, and we do not know which ones matter for combat performance.

Behavioral Parameters and System Output

Another way of asking this question of the mechanisms between "behavioral parameters/variables . . . (and) . . . system output and SOR" is to look at the three set of dimensions found in the three examples cited above and ask: How do these variables individually and collectively impact on system performance? I do not know the answer to this question, and the sources of these data do not identify the link.

One way of structuring these parameters might be in terms of the appropriate links to the basic elements of the personnel subsystem (assuming that can be defined). For example, assume that the basic elements of the personnel subsystem are:

1. Personnel selection.
2. Training.
3. Communications.
4. Command and control (leadership).
5. Tactics and doctrine.
6. Table of organization (manning).
7. Operator equipment tasks.
8. Human maintenance tasks.
9. Unit cohesiveness.
10. Mission requirements.

Presumably each of these plays a fundamental role in the performance of the personnel subsystem. And, presumably, each of the dimensions listed previously relate in some way to these parameters.

I strongly suggest that the emphasis at this time should be on the structure of the personnel subsystem, the particular measures that are related to the personnel subsystem structure, and what elements of the SOR model should reflect personnel subsystem variables. I suggest that the most important task at this time is to establish analytically (theoretically) those measures which belong in the general SOR model.

Exercise Data for Model

It is very difficult to suggest data for a model that is incompletely defined. The three examples have certainly suggested data categories, and, in one case, they have been derived from a large-sample situation.

One approach certainly could be the use of the policy-capturing technique to test the dimensions found in examples 1 and 2 for Navy operational ships. It might be of some interest to find if they are adequate for Navy ships or, if not, where they are inadequate.

The problem at this point is not getting data but, rather, getting a set of alternative models and the data they would require for testing.

Personnel Policies and Behavioral Correlates

The general question is: "How do personnel policies affect the behavior parameters/variables?" It is very difficult for me to consider this general question. I would feel much more comfortable with the development of specific examples. For example:

1. Personnel supply and personnel readiness. Suppose that, for reasons of limited personnel, supply falls short in authorized strength, either in number or skill levels. One way of demonstrating the impact of that policy is through ship simulation; we know how to do that (cf. Billingsley and Bowser, 1979) and get some rather specific answers quite quickly.

2. Training policy and skill levels in units. One possibility would be to look at specific examples of training policy and their potential impact on individual skill levels in units. Using the original model shown in equation (1), this would relate to training (T) and the need for additional training to be get the unit to an acceptable level for that dimension.

3. Retention policy and personnel readiness. We might take a look at, for example, pay policies designed for increased retention and relate them specifically to predicted MTOE required strength. This would be unit-specific and would depend upon the particular skill levels that are required for the unit.

Obviously, much data are available on this kind of correlation. Given a specific form of the model with particular dimensions defined for the personnel subsystem, it might be useful to do analytic exercises to postulate the specific kinds of relationships that might exist among specific policy decisions and the specific behavioral parameters/variables.

But, to ask this as a general question seems to me to be impossible to discuss fruitfully. Above all, we need the specific model, or alternative models, with defined parameters.

References

Billingsley, P. A., & Bowser, S. E. The simulation of ship readiness as a function of alternative manning (paper presented at ORSA/TIMS). Milwaukee: 15-17 October 1979.

Borman, W. C., & Dunnette, M. D. Selection of components to comprise a Naval Personnel Status Index (NPSI) and a strategy for investigating their relative importance. U.S. Navy: Office of Naval Research, Personnel Decisions Final Technical Report, March 1974.

Campbell, J. P., Bownas, D. A., Peterson, N. G., & Dunnette, M. D. The measurement of organizational effectiveness: A review of relevant research and opinion (Tech. Rep. 75-1). U.S. Navy: Navy Personnel Research and Development Center, July 1974. (AD-786 462)

Hawley, J. K., & Dawdy, E. D. Cost and information effectiveness analysis (CIEA): A methodology for evaluating a training device operational readiness assessment capability (DORAC) (ARI/ASA Final Rep.). U.S. Army, February 1981. (a).

Hawley, J. K., & Dawdy, E. D. Training device operational readiness assessment capability (DORAC): Feasibility and utility (ARI/ASA Final Rep.). U.S. Army, February 1981. (a)

Department of the Army. Women content in units force deployment test (MAX WAC). U.S. Army: ARI, 3 October 1977.

Department of the Army. Unit status reporting (Army Regulation (AR) 220-1). U.S. Army, 1979.

Department of the Army. The Commander's Unit Analysis Profile. U.S. Army: Ft. Hood ARI Field Unit, 1981.

DEPARTMENT OF THE ARMY
U. S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
FORT HOOD FIELD UNIT
FORT HOOD, TEXAS 76544

THE COMMANDER'S UNIT ANALYSIS PROFILE

(*Synopsis*)

The Commander's Unit Analysis Profile (CUAP) is a unit-diagnostic tool developed by the US Army Research Institute. It provides commanders of company-size units an accurate working knowledge of troop attitudes about important factors related to mission readiness and operational effectiveness.

The CUAP questionnaire, administered to E1 - E5 unit members, provides the commander a confidential estimate of the unit's standing in 21 different areas. These areas, or "factors," include such topics as: Unit Cohesiveness; Quality of Training; Officer, NCO, and Immediate Supervisor Leadership; Confidence in Unit; Alcohol/Drug-Related Problems; Race Relations; Job Satisfaction; Promotion Policy; Morale; and Reenlistment Potential.

The CUAP administration procedures require no special training and provide complete confidentiality for both the respondents and the unit commander. The questionnaire is written in easy-to-read language that can be understood by soldiers with minimal reading skills; the average soldier can complete the questionnaire in less than 20 minutes. All of the subject areas covered are those over which the unit commander can exercise control. Feedback to commanders is timely (usually within 15 days) and easily understood. The CUAP is of particular value to the newly-assigned commander who may know very little about his new unit.

Two graphical profiles, which portray the unit's scores in the 21 factor areas, are provided to the unit commander. Profile 1 depicts Unit Factor Scores, which indicate how positive or negative the unit's standing is on each factor. These scores can run from -100 (worst) to +100 (best). A score of zero would indicate "borderline." Profile 1 also shows for each factor the Average Score Other Units, which is the average score for all units in the Army that have recently utilized the CUAP. This average allows commanders to compare their units with the combined average score of all other units. Profile 2 depicts a Unit Percentile Rank for each factor. This rank is the percentage of all units that have received lower (or equal) Unit Factor Scores on the factor.

The CUAP does not replace the commander's responsibility for judging the mission readiness of the unit; rather it serves as an extremely useful tool to help isolate and identify underlying factors that may be detracting from or contributing to overall operational effectiveness.

DEPARTMENT OF THE ARMY
U. S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
FORT HOOD FIELD UNIT
FORT HOOD, TEXAS 76544

THE COMMANDER'S UNIT ANALYSIS PROFILE

> A leader need not be a psychologist, but he must have a clear valid understanding of his men and of their attitudes, aspirations, and motivations.
>
> --Army FM 22-100, "Military Leadership"

The Commander's Unit Analysis Profile (CUAP) is a unit-diagnostic tool recently developed by the US Army Research Institute. Its purpose is to provide commanders of company-size units the means for acquiring an accurate working knowledge of troop attitudes about important factors related to mission accomplishment and overall operational effectiveness.

The current 88-item CUAP questionnaire, now available for use by combat arms unit commanders, is designed to be administered to all E1 - E5 unit members. It provides the commander a completely confidential estimate of the unit's standing on each of 21 factors found to discriminate among company-size units. These factors, derived by statistical factor analysis, are described in the following paragraphs.

Factor 1. Officer Leadership: Reflects attitudes toward the overall quality of unit officer leadership, and perceptions of officer respect and concern for enlisted soldiers in the unit.

Factor 2. NCO Leadership: Reflects attitudes toward the overall quality of unit NCO leadership, and perceptions of NCO respect and concern for enlisted soldiers in the unit.

Factor 3. Immediate Supervisor Leadership: Portrays perceptions about the overall quality of supervision in the unit, supervisor treatment of and respect for the soldier-on-the-job, the clarity of supervisor job instructions and expectations, and supervisor receptiveness to constructive suggestions.

Factor 4. Leadership Concern for Soldier Welfare: Indicates the extent to which unit leadership (officers, NCOs, and immediate supervisors) is seen as having proper concern for personal and other problems of enlisted unit members.

Factor 5. Promotion Policy: Deals with the perceived fairness of and overall satisfaction with unit promotion policies, and the extent to which promotions are contingent upon performance.

[ARI,CUAP8108R273]

Factor 6. Rewards and Corrective Actions: Concerns the prevalence of positive reinforcements for jobs well-done and the treatment of "honest" mistakes and careless or intentional poor performance.

Factor 7. Leave and Pass Policies: Specifies the extent of soldier satisfaction with unit leave and pass policies.

Factor 8. Quality of Training: Represents evaluations of the quality of physical, MOS, and combat training in the unit.

Factor 9. Tools, Equipment, and Supplies: Deals with perceptions about the availability and condition of tools, equipment, and supplies.

Factor 10. Job Satisfaction: Portrays the soldiers' interest in and liking for their work, as well as their estimation of its usefulness.

Factor 11. Freedom from Harassment: Depicts opinions about the amount of harassment by superiors and the prevalence of "harassing" rules and practices in the unit.

Factor 12. Military Courtesy and Discipline: Reflects attitudes toward unit standards of discipline and courtesy, and the adherence to and enforcement of unit rules, regulations, and policies.

Factor 13. Race Relations: Describes perceptions regarding race problems and their handling in the unit and the extent to which officers, NCOs, and supervisors treat soldiers of different "races" fairly.

Factor 14. Unit Cohesiveness: Concerns interpersonal attitudes of unit members related to mutual respect, mutual confidence in one another, peer evaluation, and working together.

Factor 15. Sports Activities: Indicates the extent of soldier satisfaction with the unit sports-activities program.

Factor 16. Social Activities: Indicates the extent of soldier satisfaction with the unit social-activities program.

Factor 17. Freedom from Alcohol/Drug-Related Problems: Reveals troop perceptions of the prevalence of alcohol/drug-related problems in connection with both upper- and lower-ranking personnel in the unit.

Factor 18. Food: Indicates attitudes toward the quantity and quality of food in the field as well as in garrison.

Factor 19. Confidence in Unit: Shows the enlisted evaluation of and relative liking for the unit as a whole.

Factor 20. Morale: Portrays the soldiers' assessments of both their own morale and the morale of other soldiers in the unit.

(ARI,CUAP8108R273)

Factor 21. Reenlistment Potential: Depicts attitudes toward Army life in general and reenlistment in particular.

Two graphical unit profiles, which portray the unit's scores in the 21 factor areas, are provided to the unit commander:

1. Profile 1 depicts Unit Factor Scores, which show how positive or negative the unit's standing is on each factor. These scores can run from -100 (worst) to +100 (best). A score of zero would indicate "borderline." Profile 1 also shows for each factor the Average Score Other Units, which is the average score for all other units in the Army that have recently utilized the CUAP. This average allows commanders to compare their units with the average score of other units. Again, the scores can run from -100 to +100.

2. Profile 2 depicts a Unit Percentile Rank for each factor. This rank is the percentage of other units that have received lower (or equal) Unit Factor Scores on the factor. The percentile ranks run from 0 to 100. A rank of 75, for example, would indicate that 75 percent of other units received lower or equal scores on the factor while 25 percent received higher scores.

The CUAP administration procedures require no special training and provide complete confidentiality for both the respondents and the unit commander. The questionnaire is written in easy-to-read language that can be understood by soldiers with minimal reading skills; the average soldier can complete the questionnaire in about 20 minutes, and total administration time is usually about 30 minutes. All of the subject areas covered are those over which the unit commander can exercise control. The data are produced in a format that facilitates providing timely and easily understood feedback to commanders (usually within 15 days). The feedback permits commanders to compare their units' responses with the combined average responses of other units.

The CUAP does not replace the commander's responsibility for judging the mission readiness of the unit; rather it serves as an tool to help isolate and identify underlying factors that may be detracting from or contributing to overall operational effectiveness. The CUAP is of particular value to the newly-assigned commander who may know very little about his new unit.

The Army Research Institute's continuing research with the CUAP is devoted to making the CUAP provide as much useful information to the unit commander as possible, while requiring minimal interference with job and training time for administration. To the commander who wishes to maximize the benefit achieved from the CUAP, it is recommended that the CUAP questionnaire be administered at intervals of approximately three months (as a rule of thumb) and/or before and after crucial events the commander discerns may have a significant effect on the unit's profile.

Arrangements for administration of the CUAP questionnaire can be initiated through coordination with the Army Research Institute Field Unit, Fort Hood, Texas. POC: MAJ T. L. Pearcy, Sr. Phone: Commercial (817) 532-9826/9222/-1316; Autovon Prefix: 737.

[ARI,CUAP8108R273]

Author: Mark Sanders, California State University, Northridge

Q1. What set of behavioral (and nonbehavioral) parameters/variables would you include in the desired model and how would you define and measure these?

In considering the parameters to include, one must keep in mind the purpose to which they will be used. At a gross level, two distinct, yet interrelated, purposes can be defined. The first of which is that of model building and testing. The second is the practical purpose of assessing, or predicting, organizational effectiveness, or SOR.

Many variables we identify for building a model or testing will not be suitable for use later to assess SOR. The measures for each purpose have different criteria.

Model building/testing variables must be valid measures of the constructs we wish to tap; that is, they must possess construct validity, be relatively free of contamination, and be reliable. The practical variables, in addition, must be easy to obtain repeatedly, must be verifiable, and not subject to faking. Take, for example, "crew cohesiveness." For testing a model, we might measure it with a questionnaire, attitude scale, or sociogram. If cohesiveness, however, became part of the data collected to assess SOR, a more objective, less fakable measurement procedure would be needed (e.g., amount of time crew has performed together).

As a first stab at listing possible behavioral variables, it seems that the time-honored gross sort into individual, group, and leadership variables may be helpful. This, of course, is repeated at each level in the organization. Individual variables might break down into skill-related, attitudes, and physical health. In addition, communication between groups and across layers should be considered. Although classifying variables into categories can be valuable, it can become counterproductive if one wastes time trying to decide in which category a particular variable belongs. The following, therefore, are just a few off-the-top-of-the-head variables that might impact SOR:

1. Number of empty billets.
2. Proficiency level of the individuals (including leaders).
3. Degree of cross-training.
4. Physical condition of the individuals.
5. Experience of the individuals (including leaders) in the job position under specific mission requirements.
6. Experience of the individuals (and leaders) working as a team together.
7. Group cohesiveness.
8. Confidence in the leadership.
9. Confidence in the group's ability to perform the mission.
10. Adequacy of cross level communications.
11. Morale.
12. Respect for leader.

Q2. What system level measures would you include as being descriptive of the output of a ship/squadron/fleet and as representative of their SOR?

It seems that, to evaluate the model, SOR must be defined independent of the model parameters. If we believe SOR is related to number of personnel and availability of equipment, it would be a tautology to define SOR as a ship with a full complement of personnel and all equipment functional.

SOR should be defined, ultimately, in terms of mission performance. One could, perhaps, conceptualize a hierarchy of missions from easy/routine to difficult/complex. Further, it may not be realistic to think of SOR as the certainty of successful mission accomplishment at a given level of performance but, rather, in terms of probability of successful mission performance. Is only 100% success acceptable? Is a crew OR if there

is a 0.90 chance of successful mission performance? With a little more to drink, one could envision a scale of SOR that would allow statements to the effect that there is a 0.80 probability of successfully completing a Level 4 mission. It is doubtful, however, whether any set of parameters would allow such fine discriminations to be made.

Let me present some data I collected on organizational effectiveness (OE) measures for Army helicopter maintenance crews. It may or may not be relevant to our discussions. It does illustrate some of the complexities involved in assessing OE and some of the parameters that affect the relative importance placed on individual indices. These factors include level in the organization and wartime versus peacetime.

Three organizational levels--commanding officers (COs) (N=8), maintenance officers (MOs) (N=12), and maintenance supervisors (MSs) (N=8)--were asked to rate the importance (on a 7-pt Likert scale) of 24 indices of OE as they would relate to a helicopter maintenance crew under peacetime and wartime conditions. The 24 indices, some of which may be useful in our model, are listed below:

1. Maintenance man-hours per aircraft.
2. Maintenance man-hours per flight hour.
3. Availability.
4. Readiness.
5. Mean time to repair.
6. Down time.
7. Parts consumption.
8. Missions flown per month.
9. Flight hours per month.
10. Number of copters in for repair at any one time.
11. Number of complaints from pilots or from the unit to which the aircraft belongs.
12. Number of good parts replaced.
13. Number of work orders completed per unit time.
14. Aircraft accident rate.
15. Personnel accident rate.
16. Job satisfaction.
17. Reenlistments/turnover.
18. Morale.
19. Absenteeism.
20. Tardiness.
21. Sick call.
22. Amount of grievances.
23. Aircraft cleanliness.
24. Hangar or work area cleanliness.

The attached figures (Figures 1 and 2) show the top ten items for each level in peacetime and wartime conditions. Several things can be gleamed from the data:

1. There are differences in relative importance given to criteria depending on the level in the organization. Using all 24 indices, the correlations were:

Level	Peacetime	Wartime
CO-MO	.30	.25
CO-MS	.33	.44
MO-MS	.68	.79

2. The relative importance of the criteria change from peacetime to wartime. Correlations were .71, .57, and .38 for COs, MOs, and MSs respectively.

3. Readiness, one of the criteria, was rated fourth by all levels during peacetime, but became more important to MO and MS levels, but not to COs during wartime.

4. Those items dealing with more immediate aspects of the maintenance job tended to be rated higher by MOs and MSs than by COs.

Some questions that these data raise that may be relevant to our problem include:

1. Whether "readiness" is the appropriate criteria of organizational effectiveness, or whether other criteria, which may also reflect costs, should be factored into the picture, especially at the lower units in the organization.

2. Whether differences in the relative importance placed on various criteria by the levels of an organization might give us a hint as to the relationship between such individual variables as morale or job satisfaction and higher order criteria such as downtime, readiness, and costs? Or whether they only reflect the commanding officers' perceived likes and dislikes.

3. What impact do the differences between levels have on overall system performance? Is this a variable we should concern ourselves with in our model?

{Following the above, Dr. Sanders went directly to question 5, indicating that he was not able to address questions 3 and 4.}

Q5. <u>How do personnel policies affect the behavioral parameters/variables</u>?

One approach to this problem would be to start with a gross sort of personnel policies and regulations into potentially homogenous and useful categories. The behavioral parameters of the model could then be listed to form a two-way matrix--policies by parameters. Perhaps through logical deduction, we could assess each box of the matrix as to whether the policy might affect that parameter.

As a start, I tried to brainstorm a list of policy/regulation categories. Policies and regulations can deal with:

1. Recruitment.
2. Job placement.
3. Training.
4. Assignment to positions.
5. Organization of work (e.g., hours of work, span of control . . .).
6. Compensation.
7. Promotion.
8. Disciplinary action.
9. Retention.
10. Living conditions.
11. Recreation/leisure.

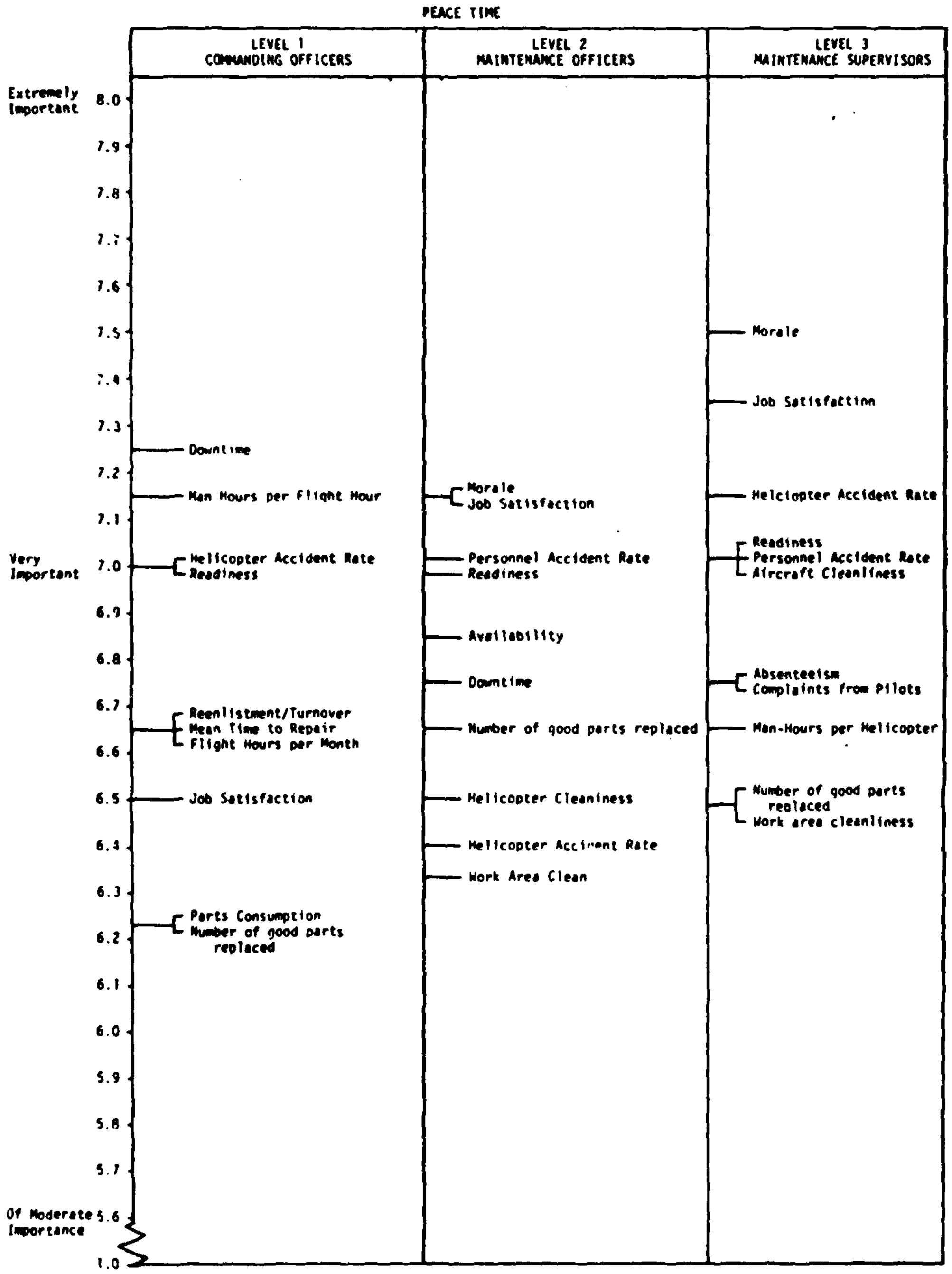


Figure 1. Rank ordering of 10 most important items rated by the three responding groups under peacetime conditions.

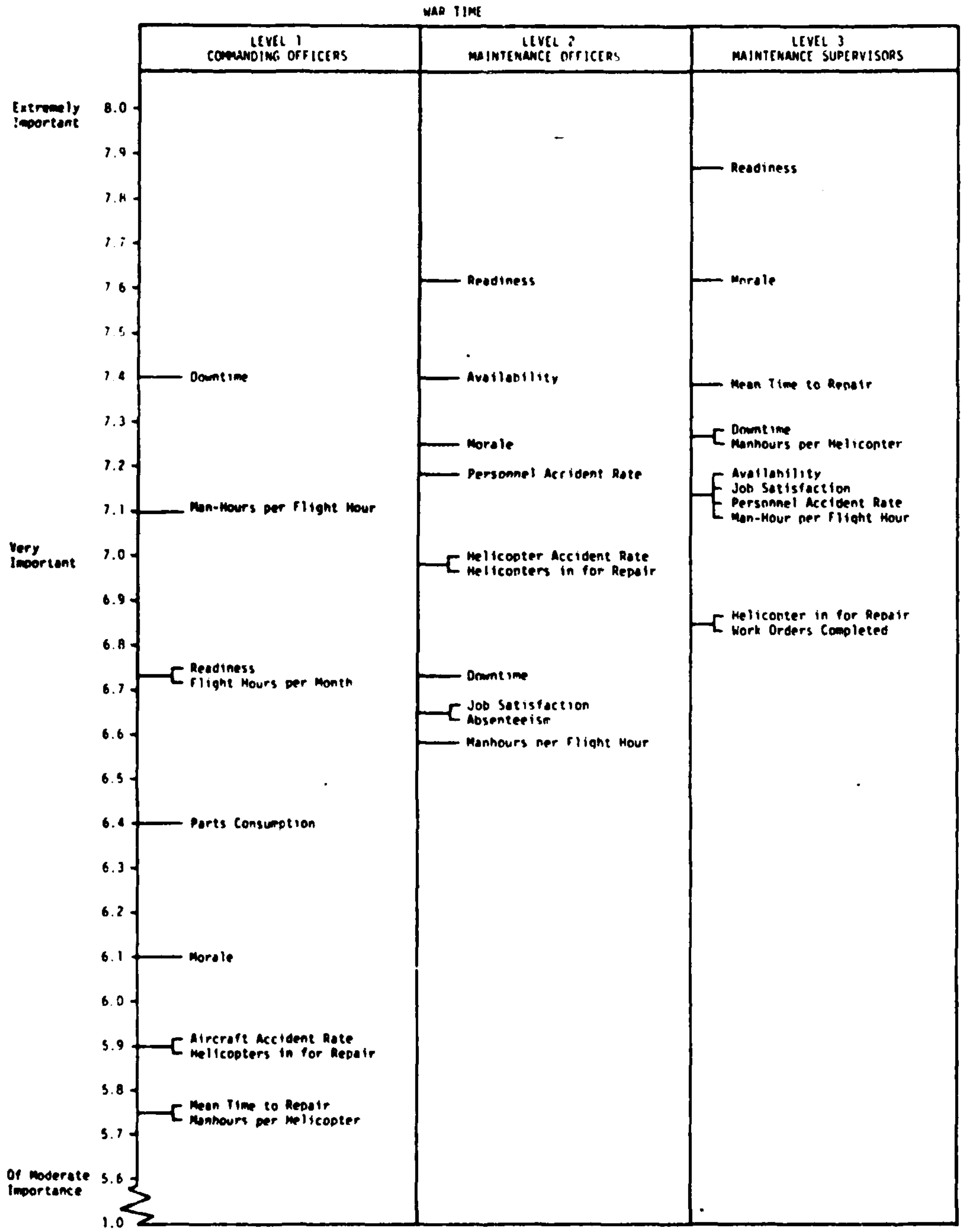


Figure 2. Rank ordering of 10 most important items rated by the three responding groups under wartime conditions.

Author: Arthur Siegel, Applied Psychological Sciences

Responses

1. What set of behavioral (and nonbehavioral) parameters/variables would you include in the desired model and how would you define and measure these?

Considering that the model is aimed at predicting readiness, it seems that the behavioral scientist should be most concerned with variables that impact "psychological readiness." Psychological readiness and the factors affecting it can probably be thought about in terms of a continuum, starting with peacetime and progressing through a preparatory period (say D-Day minus 20 to D-Day) and continuing on into the war period itself. The same or different variables may be most salient at different stages along the continuum and certainly the proportion of variance attributable to each variable may be different at different stages.

There is a host of variables that could be considered for inclusion within such a model. The problem may not be one of identifying cognitive and perceptual variables that could be salient. Rather, the problem may be that of limiting the number of variables to the most salient ones; that is, those that provide predictive variance rather than "window dressing."

Table 1 lists possible behavioral variables; and Table 2, "ground rules" for selecting variables. Which variables I would select for inclusion would depend on the answers to the "ground rule" questions vis-a-vis each variable and the type of model I wished to build--functional or psychological or mixed. Figure 1 presents my representation of each model type.

Finally, it seems that the variables built into behavioral models can be categorized as cognitive, conative, perceptual, emotional, and motor. Because psychological readiness models will depend on available behavioral theory, it is necessary to evaluate the status of the theory in each of these categories. Specifically, four questions must be answered about each category:

a. What is the adequacy of behavioral theory?

b. If competing theories are available, is it possible to make a choice among them?

c. Is the selected theory amenable to mathematical representation?

d. Are the necessary input data available or obtainable at a sufficient level of reliability/validity?

Table 3 presents a preliminary evaluation of each category relative to each question.

2. What system-level measures would you include as being descriptive of a ship/squadron/fleet and as representative of their SOR?

This question was answered, in part, above. To me, it seems that a general metric called psychological readiness is what is wanted. The psychological readiness output must be scaled so that it can serve as an input to other models which model the performance of a ship/squadron/fleet. One may ask, are not such variables as tension, conflict, expectancy, morale, cohesion, and uncertainty required outputs? I consider such variables as intervening variables which affect psychological readiness which, in turn, affects SOR.

Table 1

Examples of Possible Behavioral Variables

Information Processing	Personnel Management and Interaction
memory	self evaluation
analytic ability	evaluation of others
inductive reasoning	self confidence
deductive reasoning	social intra-extraversion
comprehension	expressional fluency
remote association	restraint
ability to see consequences	ascendence
general intelligence	sociability
evaluative ability	emotional stability
rationality	friendliness
	personal relations
Flexibility-Adaptability	leadership style
creativity	orderliness
ideational fluency	dedication to goals of installation
teamwork	dedication to goals of USA
enterprising orientation	family ties
	moral values
Personal Traits	crime involvement
carelessness	alcoholism
practical values	drug addiction
investigativeness	self control
need for autonomy	cooperation
need for change	perceived social climate
need for order	personal power
harm avoidance	persuasiveness
impulsitivity	training
Personal Characteristics	**Psychomotor Coordination**
intellective ability	speed of arm movement
training/proficiency	wrist or finger speed
	reaction time
Group Characteristics	arm-hand steadiness
cohesiveness	aiming
morale	finger dexterity
leadership	manual dexterity
communication	control precision (large movements with fine control
	multilimb coordination
	rate control
	static strength
	dynamic strength
	explosive strength
	trunk strength
	extent strength
	dynamic flexibility
	body coordination
	gross body equilibrium
	stamina
	running speed

Table 2

List of Criteria Important for Deciding Whether or Not to Include a Variable

AN ACCEPTABLE OR PREFERRED VARIABLE IS ONE...

(Data availability) which is backed by substantial empirical data for the range of population to be modeled.

(Data reliability) for which the error range of the available data is known and minimum.

(Relevance to situation being modeled) which is critical to and possesses obvious saliency to the acts and behaviors of individuals and groups involved in the situation.

(Sensitivity) which will vary as the result of various events within a situation.

(Objectivity) which can be measured in the population under consideration.

(Amenability to digital simulation) which can be modeled without unwarranted assumptions and which will not require excessive processing time or memory requirements.

(Uniqueness) which is associated with unique output variance (each variable should contribute to the richness and completeness of the total model.

(Freedom from need for artificial transformation) for which the available data do not require excessive transformation, rescaling, preprocessing, or translation for digital modeling.

(Generality) which is applicable to a range of modeled situations.

(Comprehensibility) which is easily understood by the users of the model's output.

(Utility) which is most useful for answering the questions the planner and model's user wishes to ask.

(Event-oriented) which can be updated at the conclusion of each simulated event and event indexed.

(Susceptibility to parametric variation) which can be systematically varied over a range of levels or along a continuous scale.

(Freedom from triviality) which will produce pronounced effects on input.

(Validity) for which the effects on performance (model output) and on other output and intervening variables are known, agreed on by most persons, and representative of the actual situation being modeled.

(Heuristic value) which raises questions relative to the system being simulated, as well as answer questions, provide explanations, analyze problems, provide solutions, and develop rules.

(Quantifiability) which can be quantified, measured, and scaled.

(Supportability) which is supported by a sound body of literature.

(Realism) which is apparent in the real-life situation being modeled.

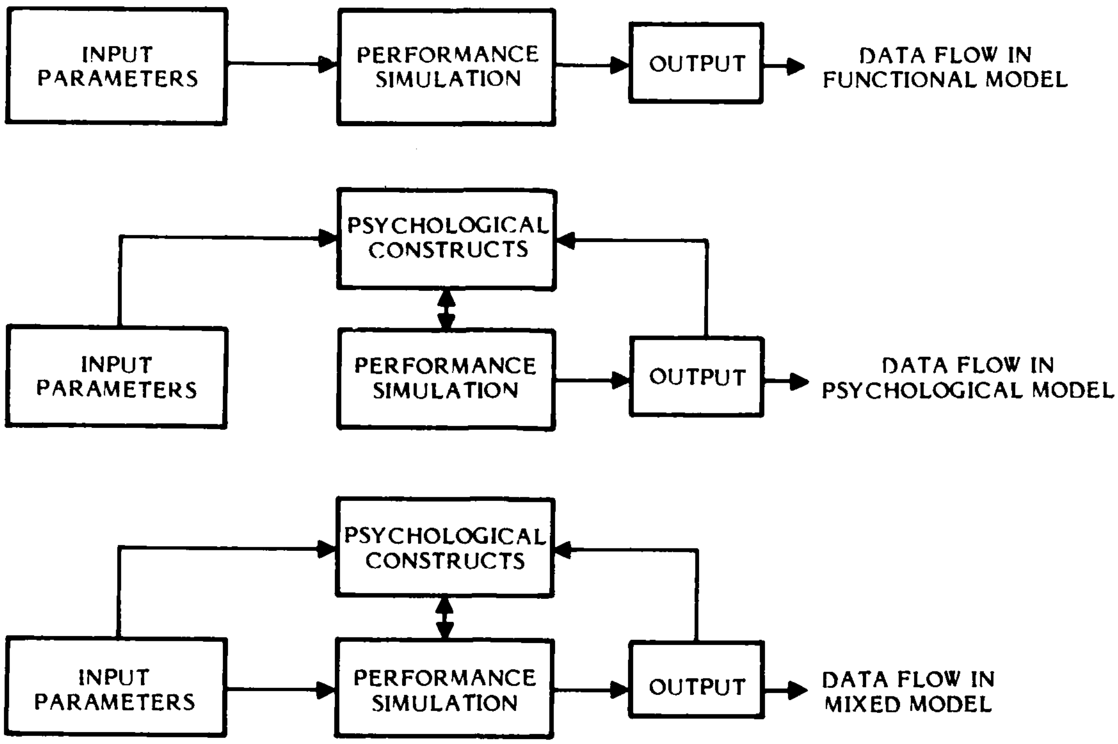


Figure 1. Data flow in functional, psychological, and mixed models.

Table 3

Evaluation of Behavioral Categories Against Four Critical Questions

	Behavioral Category				
Question	Cognitive	Conative	Perceptual	Emotional	Motor
Theory adequacy	+	0	+	0	+
Theory choice	+	0	+	+	+
Computer simulation amenability	+	0	+	0	+
Input data availability	0	-	+	-	+

Key: + = relatively high
0 = intermediate
- = relatively low

An alternate response is that the model developer should determine at the outset the questions that the model user wants answered. Knowing these, the output can be tailored to the model user's requirements. If the user wants the model to provide insight in areas that the model cannot address, then the model should probably not be developed in the first place. I do not believe that it is the modelist's responsibility to define operational readiness. This seems to be a military responsibility. The modelist's responsibility is to allow statements of the effects of varying input variables on operational readiness as defined.

3. <u>How would you conceptualize the mechanisms that permit the behavioral parameters/variables to exercise an effect on system output and SOR? Or, how does subsystem behavior link to SOR?</u>

Any model that purports to represent the wide variety of input and intervening variables that must be considered in a problem context such as that involved here will be quite complex. The input-output connections must consider the major interactions. And, there will be a large number of interactions. Accordingly, there is not a direct input-output relationship. The representations of the various relationships and interactions would depend on the specific relationships/interactions under consideration. Mathematical relationships, graph theory, stochastic simulation, multiple regression, etc., would be used as appropriate within this logic. When the set of questions came to me, I was developing a representation of the flow in a model for predicting human performance reliability in nuclear power plants. In brief, there are direct linkages, indirect linkages, and interactive linkages.

4. <u>What kind of data would one need to exercise such a model and where might one find these data?</u>

The ultimate success of a behavioral model, like any model, depends on the validity of the input data utilized. A model may be fully acceptable in terms of its constructs but may be unusable because the input data required for implementing the constructs are not available or are not at a required level of accuracy. This applies both to the numerical constants utilized as part of the model, as well as to the parameter, mission, and operator data provided as input.

While a number of data banks (e.g., Munger, Smith, & Payne)[1] are available, such data banks are remarkably barren in regard to higher order human processes and abilities. Moreover, in view of individual differences, normative data based on general populations *are often of little use to a military simulation. This input data relevancy consideration is* particularly pertinent to simulations of Navy personnel policy. Quite obviously, there is little sense in developing a sophisticated model only to find later that the required data are either meager or lacking. Too often the model designer approaches his model development/opportunities with little concern over the availability of input data. Such concerns are considered by the model developer to be mundane and are left for the model user. But, the model user wants model output results and often has little time, patience, or ability to develop or obtain esoteric, behaviorally-oriented data.

---

[1]Munger, S. J., Smith, R. W., & Payne, D. <u>An index of electronic equipment operability: Data store</u> (REP. AIR-C43-1/62-RPG). Pittsburg, PA: American Institute for Research, January 1962.

Assuming the availability of the required input data, how much burden does the model's requirements place on the user? Some models require long lists of appropriately coded input information. The input chore can be further magnified by a programmer who, with commendable computer system expertise but little human factors insight, may require features such as right justification of input or esoteric and fixed card formatting. Other model designers are more considerate, and their models are more tolerant. Input data are stored and only modifications to prestored values are required of the user; free format entry is permitted and the amount of required input data is limited.

There is obviously some upper limit on the amount of time that the user wishes to spend on preparing input data. To promote use and practicality, initial data preparation must be reasonable for initial preparation, reruns, and changes of missions and parameters. Use of interactive terminals on a conversational basis promotes this end.

The most appropriate data are obviously Navy based, war-time performance data. However, such data will (hopefully) never become available. The second order level is peacetime data (Dave Meister should jump on an airplane and fly to the Falklands to collect some performance data with the British Navy. That's the least he can do for us.) However, modelists should not hold up their work because of data quality. We should design our models to make the best use of available data while developing more and better data banks.

At Applied Psychological Services, we are not developing "data bank free" models. These work on the basis of subroutines that provide (valid) estimates of the required input data if data bank information is not available. I can expand on this approach at the workshop (if anyone except me cares about it).

5. How do personnel policies affect the behavioral parameters/variables? Or, how does one determine from an examination of a personnel policy/regulations which behavioral parameters/variables are affected by that policy/regulation? And how affected?

As a starting point, behavioral models are essentially logical representations of a process. Any logical process can be represented by a set of rules (algorithms). The question becomes one of deciding on logical relationships. This is not an armchair or intuitive process. In most areas, there is literature available that can be employed to provide such relationships. The literature may provide a theory or a set of results from which the relationships may be extrapolated. I have not yet met the case where the internal logic (content validity) cannot be developed from the literature. If the literature is barren, other options are available. For example, a conference or a Delphi technique can be employed to derive the required relationships. Quite obviously, if there is no logical, empirical, or theoretical relationship between the policy and behavioral parameters/variables, there is no basis for an algorithm and a model is not possible.

However, as noted previously, I do not believe that it is necessary, or even desirable, that a direct policy-behavioral effect exists. If this were so, there would probably be no need for a model. Operational implication questions would be easy to answer. Rather, I believe that a policy affects a number of intervening and interacting variables that, in turn, affects SOR. The beauty of a model is that it can simultaneously represent such complex relationships.

Author: H. W. Sinaiko, Smithsonian Institution

This note contains responses to the questions put to workshop participants. In addition, I am including some comments that may stimulate discussion.

Responses

Question 1. Behavioral/nonbehavioral variables for a system operational readiness (SOR) model. The traditional behavioral measures apply (i.e., speed, accuracy, and error rate of the performance of tasks and subtasks by individuals and teams). Don't overlook the long duration and sometimes continuous nature of many tasks in obtaining these measures. If an airborne ASW sonar system is supposed to be in use for a 6-hour flight, then data on its performance should be based on trials of that duration.

Not-so-traditional behavioral measures have to do with the aptitudes that people bring with them when they enter the Navy. How well do they read? Reason? Attend to complicated (or boring or simple) tasks? What is the type of supervision necessary to the operation of the system?

Many naval systems, particularly weapons, are rarely activated except during war. What do we know about the nature of skill deterioration among personnel during periods of standdown? Since few systems involve only a single operator, we need to ask about crew or team behavior: What are the effects of turbulence or personnel rotation, for example? To what extent are team members cross-trained in one another's skills?

Nonbehavioral measures should include projections of personnel supply, anticipated shortages or overages, and retention. If the Navy can expect its electronics-qualified people to remain for relatively short tours, say, through 5-6 years, there will surely be SOR consequences different from having much more experienced petty officers.

The nature of the physical Navy--equipment, software, and the like--will be critical in measuring SOR. The usual MTBF measure is an obvious approach. But, do we know anything about the effects of partial or intermittent failure of equipment components? What are the consequences, to the mission, of delay or less-than-optimum performance? Is there going to be a second chance? In this regard, I think there is a lot to be gained through innovative R&D.

Question 2. System-level measures of SOR. Again, the obvious things ae mission-related: Did we hit the target? How long did it take us to complete the task? Not so obvious, and perhaps untried, is the question of "minimal performance." How few people in the team or crew can do the job? How low can pay grades or experience be before performance is unsatisfactory? But I am assuming that most of our operational systems have some military value even if they operate at less than designed capacity. An aircraft carrier with one of its catapults inoperable is better than nothing; similarly, a ship that is manned at less than 100 percent of its complement can perform some of its assigned tasks. The question is to determine bare minima--or, conversely, to learn at what point a degraded system becomes useless or unsatisfactory.

(Because of the way the Navy seems to have developed technologically, I would put a lot of emphasis on information processing: storing, generating, transmitting, filtering, rearranging, verifying, etc. This suggests experiments on the extent to which an information system can keep command informed of casualties, trends, or other changes. The ability to deal with partial or incomplete or inaccurate information is an important

attribute. So is the ability to answer queries under time pressure. A new twist is the ability to deal with a surfeit of information--a likely requirement, in my opinion.)

System-level measures of SOR should describe how well (or poorly) the system adapts to new circumstances, such as information overload. A related question, and one that is neglected, is how well the system adapts itself to the demands of a new boss (i.e., to the idiosyncratic demands of command).

Question 3. Linking subsystem behavior to SOR. This is a hard one to deal with because every system, including the whole U.S. Navy, is a subsystem of something else; conversely, every component (almost) can be redefined as an aggregation of subsystems. My suggestion is to work in an empirical mode: Put as large a system into the laboratory as possible, run it under reasonably controllable and observable conditions, and try to fathom the relationship between major subsystem performance and the overall system.

(This is probably a good time to make the case for what Tom Belden and I have called "the indelicate experiment" (IE).[1] The method occupies a middle ground between one-time quick-and-dirty observation and fully controlled laboratory experimentation. Recently, organizational psychologists have begun to advocate "radical" (as opposed to "conservative") methods, and I think they have something like the IE in mind. In their simplest form, IEs provide a relatively low-cost, fast way to work with complex man-machine systems and to derive useful insights and answers to questions about the performance of those systems. Face validity, user involvement, and realistic tasks are key elements of IEs. The method emphasizes observation and gross measures of performance rather than tightly controlled instrumentation or voluminous data collection.)

Question 4. Data for exercising a model. Start with real-world observations of the system or a surrogate system. Don't be seduced by automatic observation techniques (e.g., cameras and tape recorders), or you'll drown in data. Determine in advance what seems to be important and go after it in the most parsimonious way. Don't overlook the value of trained observers. Critical incidents, not the Flanagan approach but the one used by the Weapon System Evaluation Group et al, in the 1960s, may be useful. But keep in mind that they can produce very sensitive (in the political sense) records. It isn't easy to reduce ethnographic observations or critical incidents to quantitative records either, but that isn't an argument against their value.

There is no single, simple pill that will solve the SOR measurement problem. Use multiple approaches, innovate, be willing to quit if an approach is ineffective. Don't collect data for purposes of making people look good. This is a sensitive issue, and if very high level people are involved, some outcomes may have to be classified. Trust and candor among researchers and the operational Navy are crucial.

Question 5. Personnel policies and behavioral variables. We don't really know, for example, how a policy limiting naval service to high school graduates will affect system operation. It seems possible that one could (and should) begin to build a list of policy effects, and, on an intuitive basis at the start, generate their presumed effects on performance. For example, if new compensation packages result in much higher career

---

[1] Sinaiko, H. W., & Belden, T. G. The indelicate experiment, in J. Spiegel and D. Walker (Eds), Information System Sciences. Washington, DC: Spartan Books, Inc., 1965.

retention rates or in large numbers of Navy men remaining in the Navy over 20 years, it would follow that we should be able to address the question of performance by older, more experienced personnel. Or, if the Navy recruits large numbers of very bright kids for short tours (e.g., 3 years), we ought to be able to predict what will happen vis-à-vis training costs, maintenance of hardware, etc. Ultimately, of course, these policy-related questions have to be answered with hard data, and it will be a long time before we can do that. The big problem now is getting started. Just listing relevant personnel policies is no mean undertaking.

Literature Available

There is a rich literature on research with large systems, most of it dating back to the 1960s and 1970s. Not much has been done since then--because, I believe, the early work was so expensive and the outcomes so meager. Nevertheless, it is important that the next generation of workers be familiar with what went before and that mistakes not be repeated. Here are some references:

1. Mac Parsons' book, Man Machine Systems Experiments, Johns Hopkins Press, 1972. The experience summarized extends to about 1967 only, but so many large and important studies were done before that time, and are included in the book, that MMSE should be required reding. The more general chapters (i.e., those not dealing with particular laboratories) may be the most pertinent today; see the chapters on methodology and on objectives and strategies. The appendices on methodology and "generalizations" are also important reading.

2. K. B. DeGreen's Systems Psychology, McGraw-Hill, 1970, contains good methodological material (e.g., Sackman's chapter on systems test and evaluation).

3. Information System Sciences, edited by Spiegel and Walker, Spartan Books, 1965 (includes the paper by Sinaiko and Belden on "the indelicate experiment").

4. Measurement of Man at Work, edited by Singleton et al., Taylor and Francis, 1979 (a broad summary of many approaches).

5. Gagnè's Psychological Principles in System Development, Holt, Rinehart and Winston, 1962, has some chapters worth re-reading (e.g., Glaser and Klaus on proficiency measurement).

Other Issues and Questions

Other issues and questions I would like to raise at the meeting:

1. Why do operators and managers traditionally resist incorporating performance into SOR and prefer the concept of SOR as a physical state (e.g., hardware, material, and human resources)?

2. Can a system's computers be used to judge the outcome of simulations and to record performance? It has been done successfully.

3. Can we exploit "soft data" for answers to hard questions? More specifically, what are history's lessons on system readiness? What was the state of the Royal Navy in April 1982 before it sailed to the South Atlantic? What did its commanders know about their fleet's readiness? There are documented cases, based on historical analysis, of Pearl Harbor--and, I suspect, of Desert I, Yom Kippur War, etc.

4. There are institutionalized systems for reporting the operational readiness of naval units. In place since 1979, a unit status and identity report (UNITREP) is a formal approach that begins with a Projected Operational Environment (POE) statement and includes, in finer detail, Required Operational Capabilities (ROCs). A POE could, for a surface ship be defined as "coastal operations in the Mediterranean"; an engineering ROC could be "four boilers operating at full capacity for 30 hours." There is a matrix in which, on one axis, ROCs are subdivided into mission areas (e.g., ASW, AAW) and quantitative ratings assigned. The other axis contains resource areas (e.g., personnel, supplies, training, and equipment) and each is rated as "fully combat ready," substantially combat ready," etc., according to preset criteria. In the behavioral or personnel area, I have been told, there is a great deal of subjectivity in establishing levels of readiness. For example, if a refrigeration specialist is called for and such a person is in fact on board, the requirement is considered fulfilled. But there is no provision for determining how many years have elapsed since the individual was trained or whether he is qualified to repair the particular equipment at hand.

Another problem with UNITREP is its failure to assess how well individuals (who may be present in the requisite numbers and skill levels) function as a team. At best the Navy can send some teams (e.g., CIC detachments) to simulator-trainers and observe their performance. But that practice is limited to a few available trainers, and it is doubtful that performance is measured with any degree of precision or that it gets into formal readiness-reporting systems. The team performance area is promising vis-a -vis behavioral aspects of SOR. Late in 1979, ONR sponsored a Rand effort to design a research problem on improving team effectiveness. A workshop was run and its outcomes published.[2] The Rand-ONR work is a significant effort, and, although aimed at training, it bears close examination in the present context. (David Meister was a participant and perhaps he can suggest a point of departure.)

5. Can techniques used to predict human performance in new systems be used to set standards of SOR? In the March 1982 issue of the ONR London European Scientific Notes, Nick Bond writes about approaches to such prediction.[3] Bond describes new applications of the microanalysis of behavior that are being developed by the Applied Psychology Unit, Admiralty Marine Technology Establishment. (Nick credits Art Siegel's work and that of Alan Swain in this domain.) My point here is not to explain the AMTE work, which I don't fully understand, but to capture its flavor and to suggest that the approach be considered as a way of measuring readiness. The AMTE method has been used to predict the performance of future submarine systems that involve many people interacting with many input-output devices. Bond characterizes these future systems as "semi-automated, computer-aided, and {involving} much data entry, updating, and deletion." AMTE employs an "activity-device matrix" (ADM) into which are entered known man-machine data (i.e., time-task and error information). The ADM axes include 31 input and output devices (e.g., keyboard, light pen, pushbutton) and about 12 behavior categories (e.g.,

---

[2] See Thorndyke, P. W., and Weiner, M. G., Improving Training and Performance of Navy Teams: A Design for a Research Program, Report R-2607-ONR, July 1980, Rand, and Goldin, S. E., and Thorndyke, P. W. (Eds.), Improving Team Performance: Proceedings of the Rand Team Performance Workshop, Report R-2606-ONR, August 1980, Rand.

[3] ESN 36-3 (1982), pp. 52-55.

selection, activation, tracking). The ADMs show, inter alia, gaps in coverage (i.e., cells where there is sufficient data) and suggest research through, say, simulation. The output of this approach seems to be useful in assessing alternative equipment configurations. (I believe it can, for our purposes, be used to develop standards or norms against which trained crews can be evaluated.) Bond points out that, in cases where alternative configurations show negligible differences, the designer can select components on the basis of cost or operator-preference criteria.

Validation studies of the AMTE method have shown that the model tends to consistently overestimate the time required to do sequences of tasks. But the method is seen as valuable for deriving rough estimates and for rank-ordering systems.

Author: Robert C. Williges, Virginia Polytechnic Institute

My preliminary comments relating to the five questions posed by David Meister represent an amplification of the comments previously distributed by Bob Mills, because I have been working with him on defining methodological considerations for systems experimentation. A complete draft of the Williges and Mills (1982) report was forwarded to Meister several weeks ago.

At some point in our discussions of SOR considerations, I feel we need to address the general questions of how to go about conducting behavioral research on large-scale systems. This issue incorporates most of the five specific questions raised by Meister in regard to behavioral parameters/variables, system-level measures, elements of a behavioral model, data requirements, and personnel policies. Specifically, I would pose that we consider methodology considerations related to three general approaches to systems research.

1. Field testing/observations. The ultimate criterion for SOR is performance in the operational environment. Short of actual combat, what type of fleet exercises can be performed to assess SOR? How will these exercises be conducted and evaluated? Techniques for field testing, field research, and quasi-experimental design (Cook & Campbell, 1979)[1] might be appropriate to consider in this regard.

2. Computer modelling. Due to the complexity of SOR considerations and the number of variables present in operational environments, computer simulation models may be quite useful in assessing SOR. Art Siegel, I am sure, will have a great deal to say about this topic, so I will not elaborate on these issues.

3. Systems experimentation. Many of the SOR considerations can be addressed through controlled systems experiments. Operator-in-the-loop simulation environments are necessary to conduct these experiments, but some of these environments currently exist and should be considered for SOR research. For example, the Advanced Command and Control Architectural Testbed (ACCAT) located in San Diego represents a testbed for investigating command and control environments.

Various methodological considerations need to be addressed in conducting these systems experiments. Parsons (1972)[2] discusses many of these topics and provides excellent examples of systems experiments. Williges and Mills (1982)[3] attempt to catalog these issues in a standard format. I attached a summary of the issues in that catalog so that other members of the workshop will have some specific examples. {See Figure 1 and Tables 1 and 2.}

References

1. Cook, T. D., & Campbell, D. T. Quasi-experimentation. Chicago: Rand-McNally, 1979.

2. Parsons, H. M. Man-machine system experiments. Baltimore: Johns Hopkins University Press, 1972.

3. Williges, R. C., & Mills, R. G. Catalog of methodological considerations for systems experimentation (Final Rep.). Wright-Patterson Air Force Base, Ohio: Aeronautical Systems Division and Air Force Aerospace Medical Research Laboratories (AFAMRL), June 1982.

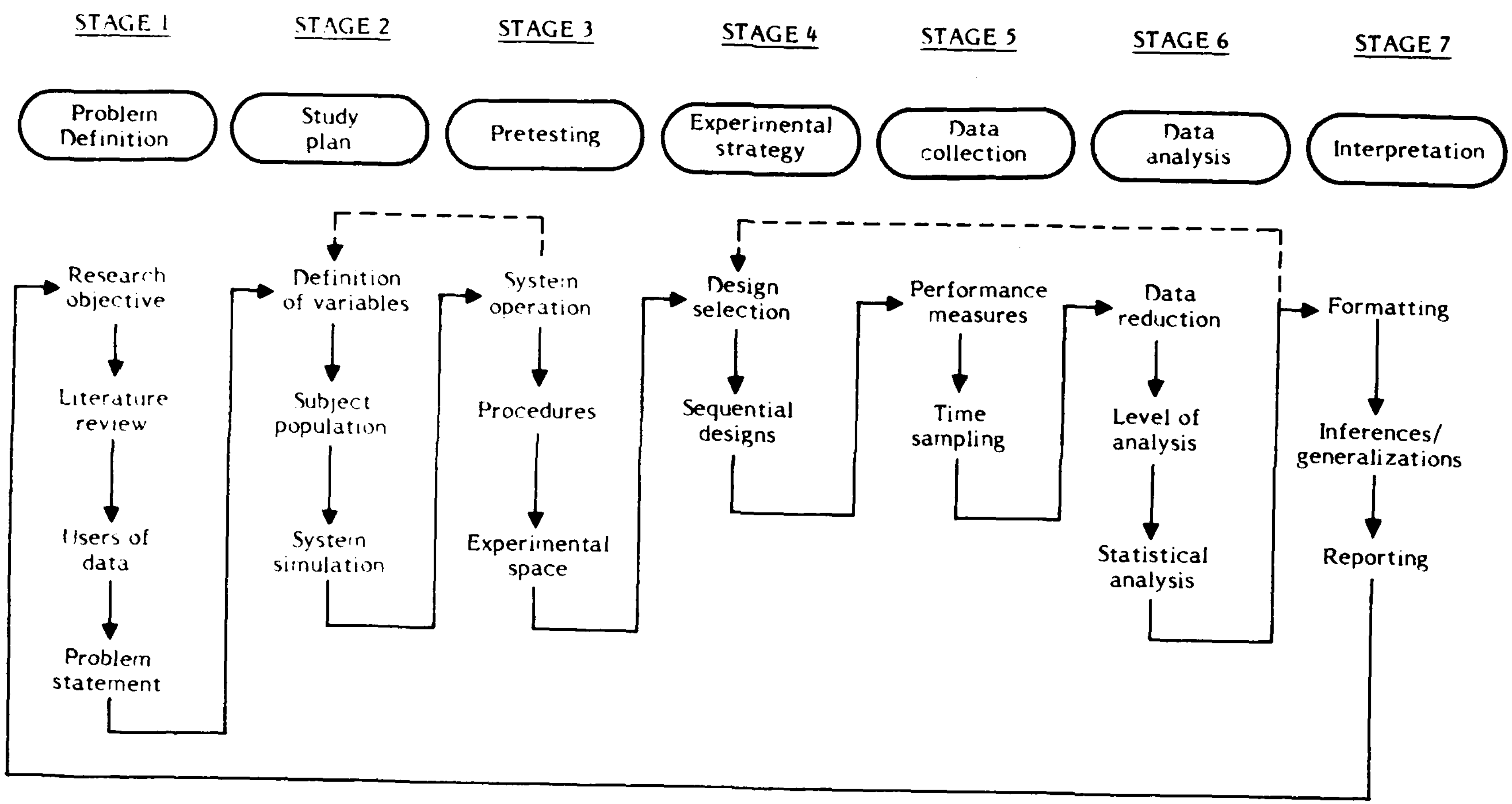


Figure 1. Flow diagram of stages involved in complex systems research.

Table 1

Classification of Issues According to Stages Involved in Complex Systems Experimentation

- 1.0 Problem Definition
  - 1.1 Required Information
  - 1.2 Research Approaches
  - 1.3 Model Building
- 2.0 Study Plan
  - 2.1 Study Plan Layout
  - 2.2 Control of Simulation Environment
  - 2.3 Quasi vs. Experimental Designs
- 3.0 Pretesting
  - 3.1 Screening Independent Variables
  - 3.2 Screening Dependent Variables
  - 3.3 Team Performance Criteria
- 4.0 Experimental Strategy
  - 4.1 Sequential Design Strategies
  - 4.2 Automated Experimental Design Aiding
  - 4.3 Qualitative vs. Quantitative Variables
- 5.0 Data Collection
  - 5.1 Multifactor Experimental Designs
  - 5.2 Assignment of Subjects
  - 5.3 Treatment of Missing Data
  - 5.4 Level of Performance Measurement
  - 5.5 Time Series Investigations
- 6.0 Data Analysis
  - 6.1 Multivariate/Multifactor Functional Relationships
  - 6.2 Estimating Optimal System Performance
  - 6.3 Colinearity/Independence of Mission Performance
  - 6.4 Pooled Data Analysis
  - 6.5 Workload Assessment
  - 6.6 Time Series Analysis
- 7.0 Interpretation
  - 7.1 Formatting and Reporting System Data

Table 2

Summary of Systems Research Methodology Issues Pertaining to Stages involved in Complex Systems Research

Issue Definition	Defining Constraints	Axioms	Research Implications
	Stage 1. Problem Definition		
1.1 Required information (See 1.2, 3.1)	Engineering design information Human performance data System vs. subsystem level	Statistical reliability Functional relationships Less than 30 variables	Develop integrated data bases Methods for conducting design tradeoffs Formal methods decomposing a system
1.2 Research approaches (See 1.1)	Literature reviews Integrated data bases Field observations Computer-analytic techniques (SAINT, HOS) Operator-in-the-loop simulation	Simulation useful in design and redesign Systems experimentation to provide engineering data Field observations used in test and evaluation	Improved field testing procedures Tradeoffs among research approaches
1.3 Model building (See 1.1, 4.1, 6.1)	Mechanistic models (Box, Hunter, & Hunter, 1978) Empirical models (Box, Hunter, & Hunter, 1978)	Normative models to describe subsystem performance Empirical models to provide functional relationships	Normative models for complete system environment Inclusion of qualitative variables Methodology for cross-referencing models with system structure
	Stage 2. Study Plan		
2.1 Study plan layout (See 3.1, 3.2, 4.1, 4.2)	Large number of factors, parameters, independent variables, constants, and dependent variables	Only one or two parameters Initial screening to reduce independent variables Constants represent operational environment	Decision rules for parameter-independent variables and constants
2.2 Control of simulation environment	Loose control Rigid control	Operational systems variables All potential variables	Methods to specify criticality of variables Tradeoffs between loose vs. rigid control
2.3 Quasi vs. experimental designs (See 4.1, 5.1, 5.5)	True experimental designs (Keppel, 1978; Winer, 1972) Quasi-experimental designs (Cook & Campbell, 1979)	Experimental designs for systems experimentation Quasi-experimental designs for field testing Interrupted time series	Guidelines for choosing appropriate designs Extension of quasi-experimental designs
	Stage 3. Pretesting		
3.1 Screening independent variables (See 2.1, 4.2)	Expert judgment vs. logical structuring Define levels based on best subjects (Geiselman & Samet, 1982) Screening studies (Simon, 1977a)	Fractional-factorial designs for screening studies 2^k vs. 3^k designs Automated screening design procedures	Effectiveness of strategies for screening studies Use of interactive automated design programs
3.2 Screening dependent variables (See 6.2)	Expert judgment Optimization procedures (Gillio, 1978) Multivariate clustering procedures (Harris, 1975)	Engineering data profiles Figure of merit Separate and composite measures Univariate and multivariate performance measures	Develop optimization procedures Evaluate clustering procedures
3.3 Team performance criteria (See 5.4, 6.3, 6.4, 6.5)	Initial team level of proficiency Roles, structure, and task allocation of teams	Allow emergence of team roles Evaluate compensatory team performance	Determine critical dimensions of teams Improve methods of team training Develop measures of compensatory team performance Develop methods for determining systems stability

Note. Modified from tables appearing in Williges & Mills, 1982. References in table are cited in Williges and Mills.

Table 2 (Continued)

Issue Definition	Defining Constraints	Axioms	Research Implications
Stage 4. Experimental Strategy			
4.1 Sequential design strategies (See 6.2)	Random selection of data (Anderson, 1953) Single-factor method (Friedman & Savage, 1947) Surface exploration (Box & Wilson, 1951) Augmented screening designs (Simon, 1973; 1977a) Strategic research (Simon, 1977b)	25% rule (Box, Hunter, & Hunter, 1978) Small experimental error (Cochran & Cox, 1958)	Comparison of alternative strategies Combining alternative strategies
4.2 Automated experimental design aiding (See 3.1, 4.1, 5.1, 6.2)	Fractional-factorial designs (Cochran & Cox, 1957) Computer-aided design (SDC, 1980; 1981) D-optimal designs (Wynn, 1970; St. John & Draper, 1975)	Knowledge of experimental designs Automated designs to determine confoundings Automated D-optimal designs for well-defined functions	Expand automated design programs Improve dialogue in automated designs Evaluate automated design procedures
4.3 Qualitative vs. quantitative variables (See 4.1, 6.1)	Quantitative variables Dummy coding qualitative variables (Draper & Smith, 1979) Semi-quantify variables	Use quantitative variables for functional relationships Interpretation of interactions with qualitative variables (Hicks, 1972)	Procedures for including qualitative variables in functional description
Stage 5. Data Collection			
5.1 Multifactor experimental designs (See 3.1, 4.1, 4.2, 6.2, 6.4)	Single-observation factorials Hierarchical designs (Winer, 1972) Blocking designs (Winer, 1972) Fractional-factorial designs (Winer, 1972) Central-composite designs (Williges, 1980)	Consider constraints in choosing design alternative Collect data in stages	Tradeoffs among various economical data collection designs Automated design aiding for various alternatives
5.2 Assignment of subjects (See 5.1)	Between-subjects, within-subjects, and mixed-factors designs (Keppel, 1978; Myers, 1972)	High efficiency of within-subjects designs Evaluate differential transfer (Poulton, 1969) Randomized blocking procedures (Keppel, 1978) Pretest to determine number of subjects	Develop automated procedures for counter-balancing within subject orderings Extend mixed-factor central composite designs
5.3 Treatment of missing data	Bias in statistical analysis (Norton, 1954; Box & Wilson, 1957) Estimate missing data Weighted and unweighted means analysis (Keppel, 1978; Winer, 1972)	Collect additional data Estimate one data point	Procedures for estimating missing data
5.4 Levels of performance measurement (See 3.3, 6.1)	Individual performance (Williges & Williges, 1982) Team performance System performance	Use multiple measures Embed performance assessment in simulation (Williges, 1979) Unconfound measures Operator vs. team-centered tasks	Use of hybrid simulation to measure team performance Develop multivariate, composite data bases Develop methods to determine set of necessary measures
5.5 Time series investigations (See 1.3, 2.3, 3.1, 3.2, 3.3, 6.5, 6.6)	Forecasting, feedback control, & intervention analysis (Box, Hunter, & Hunter, 1978) Multiple time series measures	Minimum of 50 data points (Cook & Campbell, 1979) Interrupted time series are most useful	Extend interrupted time series designs Guidelines for measuring multiple time series

<u>Note</u>. Modified from tables appearing in Williges & Mills, 1982. References in table are cited in Williges and Mills.

Table 2 (Continued)

Issue Definition	Defining Constraints	Axioms	Research Implications
	Stage 6. Data Analysis		
6.1 Multivariate/multi-factor functional relationships (See 1.3, 6.2)	Polynomial regression (Box, Hunter, & Hunter, 1978) Ridge regression (Hoerl & Kennard, 1970a; 1970b) Composite prediction equations (Williges & Williges, 1982)	Use second-order polynomials Choose orthogonal polynomial regression models No more than 7 independent variables in the final mode Collapse into multivariate clusters	Evaluate the efficacy of ridge regression Compare procedures for collapsing dependent variables Develop composite-weighting procedures for multivariate clusters Develop procedures for minimizing parameters in functional description
6.2 Estimating optimal systems performance	Partial derivatives Linear programming Brute force procedures (Gillio, 1978)	Brute force method is most useful Standard optimization requires convexity (or concavity)	Evaluations across orthogonal and nonorthogonal weights Optimization across pooled data
6.3 Colinearity/independence of mission performance (See 6.1)	Uncorrelated treatments Independent teams (Lee, 1979)	Use weighted least squares if colinearity exists (Lee, 1979)	Develop new statistical test Evaluate weighted least squares analysis
6.4 Pooled data analysis (See 5.1)	Statistical pooling procedures (Winer, 1972) Collapsing data (Williges & North, 1973) Homogeniety of regression curves (Lee, 1979) Clustering procedures (Folley & Williges, 1982) Eliminate outliers of novice subjects (Mills et al., 1981)	Use uncollapsed data analysis Collapse dependent variables into homogenous clusters Collapse across subjects to define engineering data profiles	Develop guidelines for pooling data Evaluate ridge regression of collapsed data
6.5 Workload Assessment	Behavioral measures Physiological measures (Wierwille & Williges, 1978) Team workload Peak workload (Wierwille, 1981)	Multiple measures must be used	Refine existing measures of workload Develop measures of team workload Develop measures of peak workload
6.6 Time series analysis	Theil's inequality coefficient (Kheir & Holmes, 1978) ARIMA models (Box & Jenkins, 1976)	Use Theil's inequality coefficient for multiple time series Use ARIMA models for interrupted time series	Evaluate applicability of time-series analysis
	Stage 7. Interpretation		
7.1 Formatting and reporting system data (See 3.2, 6.1)	Functional relationships System profiles (Aume & Mills, 1977) Figure of merits Retrievable data bases	Minimize statistical reports for engineering data Directly relate to real system configuration	Compare methods for formatting data Computer augmentation of system data profiles

Note. Modified from tables appearing in Williges & Mills, 1982. References in table are cited in Williges and Mills.

DISTRIBUTION LIST

Chief of Naval Operations (OP-112), (OP-115) (2), (OP-987H)
Chief of Naval Research (Code 440) (3)
Chief of Naval Education and Training (N-21)
Chief, Army Research Institute Field Unit--USAREUR (Library)
Commander, Air Force Human Resources Laboratory, Brooks Air Force Base (Manpower and Personnel Division)
Commander, Air Force Human Resources Laboratory, Brooks Air Force Base (Scientific and Technical Information Office)
Commander, Air Force Human Resources Laboratory, Lowry Air Force Base (Technical Training Branch)
Commander, Air Force Human Resources Laboratory, Williams Air Force Base (AFHRL/OT)
Defense Technical Information Office (DDA) (12)

END

FILMED

12-83

DTIC